Benutzerhandbuch

StatTools

Statistik-Add-In für Microsoft® Excel

Version 7
Juni, 2015

Palisade Corporation
798 Cascadilla Street
Ithaca, NY 14850
USA
+1-607-277-8000
+1-607-277-8001 (Fax)
http://www.palisade.com (Web-Site)
sales@palisade.com (E-Mail)

Copyright-Hinweis
Copyright © 2015, Palisade Corporation.

Warenzeichen
Microsoft, Excel und Windows sind eingetragene Warenzeichen der Microsoft Corporation.
IBM ist ein eingetragenes Warenzeichen von International Business Machines, Inc.
Palisade, TopRank, BestFit und RISKview sind eingetragene Warenzeichen der Palisade Corporation.

Willkommen in StatTools für Excel

Willkommen

Durch StatTools erhält Microsoft Excel, das den Industrienormen entsprechende Daten- und Analysen-Modellierungswerkzeug, einen neuen und leistungsfähigen Werkzeugsatz zum Erstellen von Statistiken! StatTools ist ein Add-In für die Statistikerstellung in Microsoft Excel. Dadurch können Sie in Excel-Arbeitsblättern Daten analysieren, während Sie in der Ihnen bekannten Microsoft Office-Umgebung arbeiten. Durch das Kombinieren einer leistungsfähigen Datenverwaltung mit Analysen, die sich mit den besten auf dem Markt befindlichen Statistikpaketen messen können, bietet StatTools Ihnen beiderseits das beste, das heißt mühelose, Microsoft Office entsprechende Verwendung und Berichte sowie auch robuste statistische Leistung.

Arbeiten Sie bequem

Wenn Sie mit Excel vertraut sind, kennen Sie auch StatTools! StatTools funktioniert genau wie Excel, d.h. mit Symbolleisten, Menüs und benutzerdefinierten Arbeitsblattfunktionen – und diese alles direkt in Excel. Im Gegensatz zu eigenständiger Statistiksoftware gibt es bei StatTools keine steile Lernkurve und auch keine im voraus zu zahlenden Schulungskosten, da Sie wie gewohnt direkt in Excel arbeiten. Ihre Daten und Variablen befinden sich in den Excel-Kalkulationstabellen. Mit anderen Worten, Sie können die standardmäßigen Excel-Formeln für Berechnungen und Umformungen verwenden sowie auch Excel-Sortierung und Pivot-Tabellen. Die sich aus Ihren Statistikanalysen ergebenden Berichte und Diagramme sind im standardmäßigen Excel-Format und können in Zusammenhang mit allen in Excel integrierten Formatierungsfähigkeiten verwendet werden.

Robuste Statistiken in Excel

StatTools ersetzt die in Excel integrierte Statistik durch seine eigenen robusten und schnellen Berechnungen. Die Genauigkeit der in Excel verwendeten Statistiken ist schon oft in Frage gestellt worden und daher werden diese in StatTools überhaupt nicht eingesetzt! Sogar die in Excel vorhandenen Arbeitsblattstatistikfunktionen, wie z.B. *STABW()*, sind durch neue, robuste StatTools-Versionen, wie z.B. *StatSTDEV()*, ersetzt worden. Die in StatTools verwendeten Statistikberechnungen sind äußerst genau und leistungsoptimiert durch Verwendung von C++ DLLs anstelle von Makro-Berechnungen.

StatTools-Analysen

Durch StatTools können die gängigsten Statistikverfahren ausgeführt werden. Auch bietet StatTools beispiellose Fähigkeiten zum Hinzufügen von neuen benutzerdefinierten Analysen. Mithilfe von insgesamt 36 weitreichenden statistischen Vorgängen plus 8 integrierten Daten-Dienstprogrammen sind in StatTools fast alle allgemein verwendeten Statistikanalysen verfügbar. Diese Statistikfunktionen schließen beschreibende Statistiken, Normalitätstests, Gruppenvergleiche, Korrelation, Regressionsanalyse, Qualitätskontrolle, Prognosen und auch noch anderen Funktionen mit ein. Hinzu kommen noch die vielen benutzerdefinierten Vorgänge (die von Ihren eigenen Mitarbeitern oder anderen Fachleuten geschrieben wurden). Mit anderen Worten, durch StatTools steht Ihnen direkt in Excel ein umfassender und anpassbarer Werkzeugsatz zum Erstellen von Statistiken zur Verfügung.

StatTools bietet Ihnen in Echtzeit automatisch aktualisierte Statistikberechnungen! Wenn Sie in Excel einen Wert ändern, erwarten Sie, dass Ihr Arbeitsblatt neu berechnet wird und Ihnen ein neues Ergebnis zeigt. Das gleiche geschieht praktisch in StatTools! Wenn Sie einen Wert in Ihrem Datensatz ändern, wird Ihr Statistikbericht automatisch aktualisiert. StatTools verwendet einen leistungsstarken Satz von benutzerdefinierten Arbeitsblattfunktionen, um sicherzustellen, dass die in Ihren Berichten erscheinenden Statistiken immer den aktuellen Daten entsprechen.

StatTools-Datenverwaltung

StatTools bietet Ihnen einen umfassenden Datensatz und auch die entsprechende Datenverwaltung direkt in Excel, genauso wie Sie es von einem eigenständigen Statistikpaket erwarten würden. Sie können direkt von den Daten in Excel aus die Anzahl der Datensätze nebst Variablen definieren, die analysiert werden sollen. StatTools greift auf intelligente Weise auf Ihre Datenblöcke zu und schlägt entsprechende Namen für Variablen und Speicherorte für Daten vor. Ihre Datensätze und Variablen können sich in verschiedenen Arbeitsmappen und Arbeitsblättern befinden. Mit anderen Worten, Sie können Ihre Daten wie gewünscht einrichten. Anschließend führen Sie Statistikanalysen aus, die auf Ihre Variablen verweisen, damit die Daten in Excel nicht immer wieder erneut ausgewählt werden müssen. In StatTools sind die Variablen außerdem auch nicht nur auf eine einzelne Datenspalte in einem Excel-Arbeitsblatt beschränkt. Sie können für eine Variable dieselbe Spalte in mehreren Arbeitsblättern verwenden! In Excel 2007 und späteren Versionen sind mehr als ein Million Zeilen in einem Arbeitsblatt vorhanden. Auch ist bei diesen Versionen die Anzahl der Arbeitsblätter in einer Arbeitsmappe nicht beschränkt. Das bedeutet, dass in *StatTools Industrial* die Anzahl der Testfälle, die analysiert werden können, nur durch den verfügbaren Speicherplatz begrenzt ist. StatTools Professional ist dagegen auf 10.000 Testfälle beschränkt.

StatTools-Berichte

Excel kann besonders gut für Berichte und Diagramme verwendet werden und StatTools macht sich das sehr zu Nutze. In StatTools werden Diagramme im Excel-Format verwendet, die mühelos neuen Farben, Schriftarten und zusätzlichem Text angepasst werden können. Berichtstitel, Zahlenformate und Texte können genau wie in jedem standardmäßigen Excel-Arbeitsblatt mühelos geändert werden. Auch können Sie Tabellen und Diagramme aus den StatTools-Berichten direkt in Ihre in anderen Anwendungen befindlichen Dokumente gezogen und dort abgelegt werden. Diagramme und Tabellen bleiben mit den Daten in Excel verknüpft. Mit anderen Worten, bei Änderung Ihrer Analyseberichte wird Ihr Dokument automatisch entsprechend aktualisiert.

Datenzugriff und -freigabe

Excel bietet ausgezeichnete Importfunktionen. Ihre bereits vorhandenen Daten können daher mühelos in StatTools importiert werden! Über die standardmäßigen Excel-Funktionen können daher Daten aus Microsoft SQL Server, Oracle, Microsoft Access oder irgendeiner anderen mit ODBC konform gehenden Datenbank eingelesen werden. Daten können aus Textdateien oder auch anderen Anwendungen geladen werden. Mit anderen Worten, solange ein Einlesen der Daten in Excel möglich ist, können sie auch in StatTools verwendet werden!

Alle Ergebnisse und Daten aus StatTools werden in Excel-Arbeitsmappen gespeichert. Sie können Ihre Ergebnisse und Daten aus StatTools direkt an andere Kollegen oder Mitarbeiter senden, genauso wie das bei anderen Excel-Dateien der Fall ist. Freigeben ist einfach Kinderspiel!

Inhaltsverzeichnis

Willkommen in StatTools für Excel i

Willkommen i

Arbeiten Sie bequem i
Robuste Statistiken in Excel ii
StatTools-Analysen ii
StatTools-Datenverwaltung iii
StatTools-Berichte iii
Datenzugriff und -freigabe iv

Inhaltsverzeichnis v

Kapitel 1: Erste Schritte 1

Einführung 3

Info zu dieser Version 3
Die Betriebssystemumgebung 3
Unterstützung 3
Systemanforderungen von StatTools 6

Installationsanleitung 7

Allgemeine Installationsanleitung 7
Konfiguration der StatTools-Symbole oder Verknüpfungen 7
DecisionTools Suite 8

Softwareaktivierung 9

Kapitel 2: Überblick über StatTools 11

Überblick 13

StatTools-Menü und StatTools-Symbolleiste 13
Datensätze und Datenverwaltung 13
Berichte und Diagramme in StatTools 16

Kapitel 3: StatTools-Referenzhandbuch 19

Einführung 23

VBA-Makrosprache und Toolkit für Entwickler in StatTools 23

Referenz: StatTools-Symbole 25

StatTools-Symbolleiste 25

Referenz: StatTools-Menübefehle 27

Einführung 27

Auflistung der Befehle 28

StatTools-Menü – Datensätze 95

Befehl *Datensatzmanager* 95

Menü *Dienstprogramme* 101

Befehl *Stapeln* 101
Befehl *Entstapeln* 104
Befehl *Daten transformieren* 105
Befehl *Verzögerung* 107
Befehl *Differenz* 109
Befehl *Interaktion* 111
Befehl *Kombination* 113
Befehl *Pseudo* 115
Befehl *Zufallsprobe* 117

Menü *Übersichtsstatistik* 119

Befehl *Übersicht (eine Variable)* 119
Befehl *Korrelationen und Kovarianz* 122

Menü *Übersichtsdiagramme* 125

Befehl *Histogramm* 125
Befehl *Punktdiagramm* 128
Befehl *Box-Whisker-Plot* 131

Menü *Statistischer Schluss* 135

Befehl *Vertrauensintervall – Mittelwert/Std. Abweichung* 135
Befehl *Vertrauensintervall für Proportionen* 138
Befehl *Hypothesetest für Mittelwert/Std. Abweichung* 141
Befehl *Hypothesetest für Proportion* 144
Befehl *Probegrößenauswahl* 148
Befehl *Einseitige ANOVA* 150
Befehl *Zweiseitige ANOVA* 154
Befehl *Chi-Quadrat-Unabhängigkeitstest* 157

Menü *Normalitätstests* 161

Befehl *Chi-Quadrat-Normungstest* 161
Befehl *Lilliefors-Test* 165

Befehl *Normales Q-Q-Diagramm* **167**

Menü *Zeitserie und Prognose* **169**

Befehl *Zeitserien-Diagramm* **169**
Befehl *Autokorrelation* **172**
Befehl *Ausführzählungstest für Zufallsfaktoren* **174**
Befehl *Prognose* **176**

Menü *Regression und Klassifizierung* **181**

Befehl Regression **181**
Befehl *Logistische Regression* **193**
Befehl *Diskriminanzanalyse* **198**

Menü *Qualitätskontrolle* **203**

Befehl *Pareto-Diagramm* **204**
Befehl *x/r-Diagramme* **207**
Befehl *p-Diagramm* **211**
Befehl *c-Diagramm* **215**
Befehl *u-Diagramm* **218**

Menü *Nichtparametrische Tests* **223**

Befehl *Vorzeichentest* **225**
Befehl *Wilcoxon-Vorzeichen-Rangtest* **228**
Befehl *Mann-Whitney-Test* **231**
Befehl Kruskal-Wallis-Test **235**

Menü *multivariate Analyse* **239**

Befehl Hauptkomponenten-Analyse **239**
Befehl Clusteranalyse **243**

Menü *Dienstprogramme* **251**

Befehl *Anwendungseinstellungen* **251**
Befehl *Datensätze löschen* **258**
Befehl *Dialogfeldspeicher löschen* **258**
Befehl *StatTools-Add-In entladen* **258**

Hilfemenü **259**

StatTools-Hilfe **259**
Befehl *Online-Handbuch* **259**
Befehl *Lizenzaktivierung* **259**
Befehl *Info über* **259**

Referenz: StatTools-Funktionen **261**

Einführung **261**

StatTools-Funktionen gegenüber Excel-Funktionen **261**

Verteilungsfunktionen **263**
Echtzeit-Berichte **264**

Referenz: Liste der Statistikfunktionen **267**

Tabelle der verfügbaren Funktionen **267**
Detaillierte Funktionsbeschreibungen **270**

Index Error! Bookmark not defined.

Kapitel 1: Erste Schritte

Einführung **3**

Info zu dieser Version **3**
Die Betriebssystemumgebung **3**
Unterstützung **3**
Systemanforderungen von StatTools **6**

Installationsanleitung **7**

Allgemeine Installationsanleitung **7**
Konfiguration der StatTools-Symbole oder Verknüpfungen **7**
DecisionTools Suite **8**

Softwareaktivierung **9**

Einführung

In dieser Einführung wird der Inhalt des StatTools-Paketes beschrieben und Ihnen gezeigt, wie StatTools installiert und mit Microsoft Excel verknüpft werden kann.

Info zu dieser Version

Diese Version von StatTools kann zusammen mit Microsoft Excel 2007 oder höher verwendet werden.

Die Betriebssystemumgebung

Dieses Benutzerhandbuch geht davon aus, dass Sie allgemein mit dem Windows-Betriebssystem und mit Excel vertraut sind. Das heißt, es wird angenommen:

- *dass Sie sich mit dem Computer und der Maus auskennen*
- *dass Ihnen Begriffe wie Symbol, Klicken, Doppelklicken, Menü, Fenster, Befehl und Objekt bekannt sind*
- *dass Sie grundlegende Konzepte wie „Verzeichnisstruktur" und „Dateibenennung" verstehen*

Unterstützung

Allen registrierten StatTools-Benutzern mit gültigem Wartungsplan steht unser technischer Support kostenlos zur Verfügung. Benutzer ohne Wartungsplan können unseren technischen Support gegen Berechnung per Vorfall in Anspruch nehmen. Um sicherzustellen, dass Sie als StatTools-Benutzer registriert sind, **sollten Sie die Registrierung online über unsere Website http://www.palisade.com/support/register.asp** vornehmen.

Wenn Sie sich telefonisch mit uns in Verbindung setzen, sollten Sie immer die Seriennummer und das Benutzerhandbuch parat haben. Außerdem können wir Sie technisch besser unterstützen, wenn Sie vor dem Computer sitzen und arbeitsbereit sind.

Bevor Sie anrufen…

Bevor Sie unseren technischen Support anrufen, ist es angebracht, folgende Prüfliste nochmals abzuhaken:

- *Haben Sie sich die Online-Hilfe angesehen?*
- *Haben Sie in diesem Benutzerhandbuch nachgeschlagen und auch das Multimedia-Lernprogramm online durchgearbeitet?*
- *Haben Sie die Datei README.WRI gelesen? Sie enthält aktuelle StatTools-Informationen, die evtl. bei Drucklegung des Handbuchs noch nicht zur Verfügung standen.*
- *Können Sie das Problem nachvollziehen? Kann das Problem auch auf einem anderen Computer oder bei einem anderen Modell nachvollzogen werden?*
- *Haben Sie sich bereits unsere Web-Seite* ***(http://www.palisade.com)*** *angesehen? Sie enthält die neueste FAQ (eine durchsuchbare Datenbank mit Fragen und Antworten, welche den technischen Support betreffen) sowie StatTools-Patches (Korrekturprogramme), die unter „Technical Support" zu finden sind. Wir empfehlen Ihnen, regelmäßig unsere Web-Seite aufzusuchen, damit Sie sich laufend über die neuesten StatTools-Informationen sowie über anderweitige Palisade-Software informiert halten können.*

Kontaktieren von Palisade

Palisade Corporation ist dankbar für alle Fragen, Bemerkungen oder Vorschläge, die mit StatTools zu tun haben. Es gibt viele Möglichkeiten, sich mit unserer technischen Abteilung in Verbindung zu setzen, zum Beispiel:

- *per E-Mail:* ***support@palisade.com***
- *per Telefon: +1-607- 277-8000 (montags bis freitags zwischen 9.00 und 17.00 Uhr US-Ostküstenzeit). Lassen Sie sich zum „Technical Support" durchschalten*
- *per Fax: +1-607-277-8001*
- *per Post:*

  Technischer Support
  Palisade Corporation
  798 Cascadilla St.
  Ithaca, NY 14850, USA

Palisade Europe ist wie folgt zu erreichen:

- *per E-Mail:* ***support@palisade-europe.com***
- *per Telefon: +44 1895 425050 (GB)*
- *per Fax: +44 1895 425051 (GB)*
- *per Post:*

  Palisade Europe
  31 The Green
  West Drayton
  Middlesex
  UB7 7PN
  Großbritannien

Palisade Asia Pacific ist wie folgt zu erreichen:

- *per E-Mail:* ***support@palisade.com.au***
- *per Telefon: +61 2 9252 5922 (AU)*
- *per Fax: +61 2 9252 2820 (AU)*
- *per Post:*

  Palisade Asia-Pacific Pty Limited
  Suite 404, Level 4
  20 Loftus Street
  Sydney NSW 2000
  Australien

Es ist wichtig, dass Sie uns bei jeder Kommunikation den Produktnamen, die Version sowie die Seriennummer nennen. Sie können die Versionsnummer herausfinden, indem Sie in Excel im StatTools-Menü auf **Hilfe über** klicken.

Versionen für Studenten

Für die Studentenversion von StatTools steht kein telefonischer Support zur Verfügung. Wenn Sie bei dieser Version Hilfe benötigen, sollten Sie eine der folgenden Alternativen versuchen:

- *fragen Sie Ihren Professor bzw. Lehrbeauftragten*
- *sehen Sie auf unserer Website http://www.palisade.com unter „Answers to Frequently Asked Questions" (Antworten auf häufig gestellte Fragen) nach*
- *wenden Sie sich per E-Mail oder Fax an unsere Abteilung „Technical Support"*

Systemanforderungen von StatTools

Bei StatTools für Microsoft Excel für Windows sind folgende Systemanforderungen zu berücksichtigen:

- *Microsoft Windows XP oder höher*
- *Microsoft Excel 2007 oder höher*

Installationsanleitung

Allgemeine Installationsanleitung

Durch das Setup-Programm werden die StatTools-Systemdateien in das Verzeichnis kopiert, das Sie auf der Festplatte angegeben haben. So wird das Setup-Programm unter Windows XP oder höher ausgeführt:

1) *Doppelklicken Sie beim Ausführen der Installations-CD auf StatTools Setup.exe und folgen Sie dann den Installationsanweisungen auf dem Bildschirm.*

Falls Sie bei der Installation von StatTools auf Probleme stoßen, sollten Sie nachsehen, ob genügend Speicherplatz auf dem Laufwerk verfügbar ist, auf dem TopRank installiert werden soll. Versuchen Sie dann die Installation erneut, nachdem Sie ausreichend Speicherplatz freigemacht haben.

Deinstallation von StatTools

Wenn Sie StatTools dagegen entfernen möchten, müssen Sie das Dienstprogramm **Software** in der Systemsteuerung verwenden und dann den Eintrag **StatTools** auswählen.

Konfiguration der StatTools-Symbole oder Verknüpfungen

Erstellung der Verknüpfung in der Windows Task-Leiste

In Windows wird durch das Setup-Programm automatisch ein StatTools-Befehl (Start\Programme\Palisade DecisionTools) in der Task-Leiste erstellt. Sollten jedoch während der Installation Probleme auftreten, oder aber wenn Sie das Konfigurieren der Programmgruppe und Symbole zu einer anderen Zeit manuell vornehmen möchten, gehen Sie bitte wie folgt vor. Bei den nachstehenden Gebrauchsanweisungen wird von Windows XP Professional ausgegangen. Bei anderen Betriebssystemen muss evtl. etwas anders vorgegangen werden.

1) *Klicken Sie auf* ***Start*** *und zeigen Sie dann auf* ***Einstellungen****.*

2) *Klicken Sie auf* ***Task-Leiste*** *und* ***Startmenü*** *und anschließend auf die Registerkarte* ***Startmenü****.*

3) *Klicken Sie auf* ***Anpassen****, dann auf* ***Hinzufügen*** *und danach auf* ***Durchsuchen****.*

4) *Suchen Sie nach der Datei* ***StatTools.exe,*** *klicken Sie auf diese Datei und anschließend auf* ***OK****.*

5) *Klicken Sie auf **Weiter** und doppelklicken Sie anschließend auf das Menü, in dem das Programm erscheinen soll.*

6) *Geben Sie den Namen **StatTools** ein und klicken Sie schließlich auf **Beenden**.*

7) *Klicken Sie in allen geöffneten Dialogfeldern auf **OK**.*

DecisionTools Suite

StatTools ist eine Komponente der DecisionTools Suite, bei der es sich um einen Satz von Produkten für die Risiko- und Entscheidungsanalyse handelt, die von Palisade Corporation erhältlich sind. Normalerweise wird StatTools in einem Unterverzeichnis von *Programme\Palisade* installiert. Das ist so ähnlich, wie z. B. Excel oft in einem Unterverzeichnis von **Microsoft Office** installiert wird.

Eines der Unterverzeichnisse von *Programme\Palisade* ist somit das StatTools-Verzeichnis, das gewöhnlich die Bezeichnung **Stattools7** hat. Dieses Verzeichnis enthält dann die StatTools-Add-In-Programmdatei (STATTOOLS.XLA) sowie auch Beispielmodelle und andere zur Ausführung von StatTools erforderliche Dateien. Ein anderes Unterverzeichnis von *Programme\Palisade* ist das Verzeichnis SYSTEM, in dem sich die Dateien befinden, die von den einzelnen Programmen der DecisionTools Suite benötigt werden (einschließlich Hilfedateien und Programmbibliotheken).

Softwareaktivierung

Bei der Aktivierung handelt es sich um einen einmaligen Lizenzprüfprozess, der erforderlich ist, um die Palisade-Software als vollkommen lizenziertes Produkt ausführen zu können. Die **Aktivierungs-ID** befindet sich auf der an Sie geschickten Rechnung und sieht so ähnlich wie z. B. „DNA-6438907-651282-CDM“ aus. Wenn Sie diese Aktivierungs-ID während der Installation eingeben, ist die Software bereits bei Beendung des Installationsvorgangs aktiviert und daher keine weitere Aktivierung durch Sie erforderlich. Falls Sie die Software erst irgendwann nach der Installation aktivieren möchten, müssen Sie im Hilfemenü den Befehl **Lizenzmanager** wählen.

Mithilfe des Lizenzmanagers können Software-Lizenzen aktiviert, deaktiviert und auch auf einen anderen Computer verlegt werden. Ebenfalls kann der Lizenzmanager zum Verwalten von Lizenzen für Netzwerkinstallationen verwendet werden. Folgen Sie im Lizenzmanager dann den entsprechenden Anweisungen und Dialogen, um den gewünschten Lizenzierungsvorgang auszuführen.

Kapitel 2: Überblick über StatTools

Überblick **13**

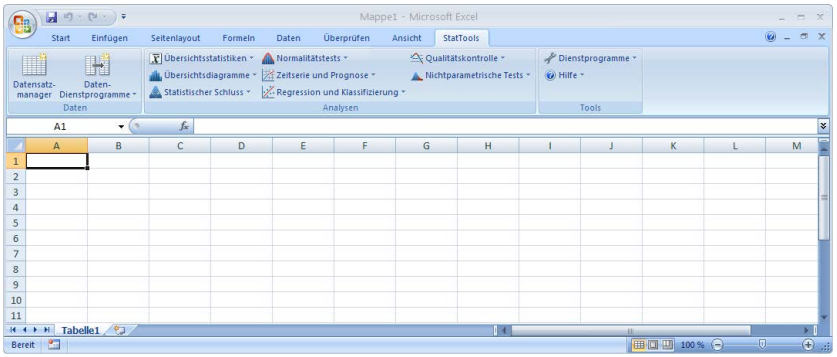
StatTools-Menü und StatTools-Symbolleiste **13**
Datensätze und Datenverwaltung **13**
Berichte und Diagramme in StatTools **16**

Überblick

StatTools gibt Ihnen leistungsstarke Statistikfähigkeiten in einer Ihnen bekannten Umgebung – dem Microsoft Excel-Programm. StatTools-Vorgänge, wie z.B. Erstellen von Punktdiagrammen, Normungstests für Variablen und Ausführen von Regressionsanalysen, können an Ihren Daten direkt in Excel vorgenommen werden und die sich aus den Analysen ergebenden Berichte und Diagramme werden ebenfalls in Excel erstellt.

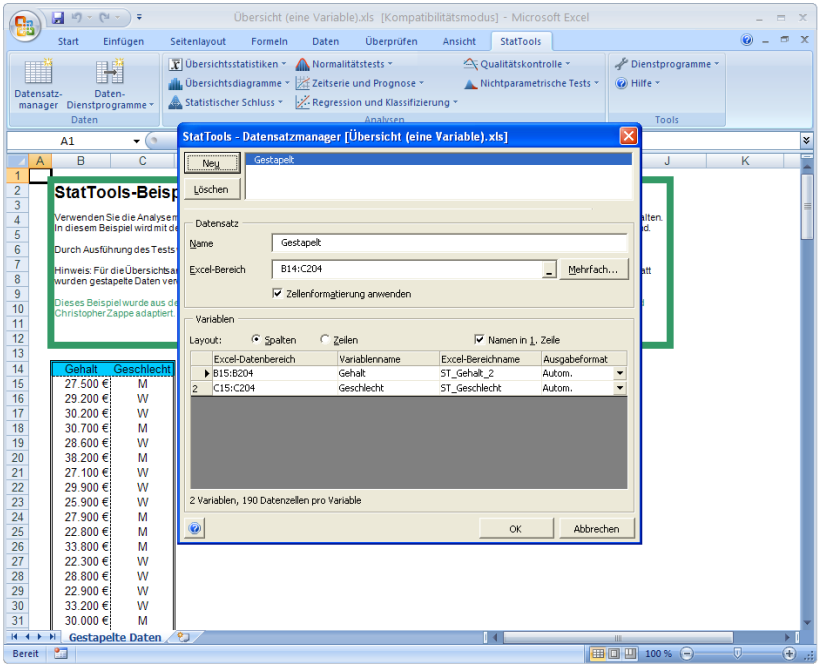
StatTools-Menü und StatTools-Symbolleiste

In Excel 2007 und späteren Excel-Versionen wird dagegen eine StatTools-Befehlsleiste angezeigt.



Datensätze und Datenverwaltung

StatTools funktioniert so ähnlich wie die meisten eigenständigen statistischen Softwarepakete, indem es den Variablen gemäß strukturiert ist. Bei den meisten Analysen ist es erforderlich, dass Sie mit einem Datensatz oder einem Satz von statistischen Variablen arbeiten, die sich oft in zusammenhängenden Spalten mit verschiedenen Namen befinden. Die Namen der Variablen sind gewöhnlich in der ersten Zeile des Datensatzes zu sehen. In StatTools ermöglicht Ihnen der **Datensatzmanager** jedoch, die gewünschten Datensätze und Variablen zu definieren. Diese vordefinierten Variablen können dann in den Statistikanalysen verwendet werden, und zwar ohne dass die zu analysierenden Daten immer wieder neu ausgewählt werden müssen.



Jede in einem Datensatz befindliche Variable hat einen Namen und ist in Excel mit einem bestimmten Zellenbereich verknüpft. Gewöhnlich ist **eine Variable pro Spalte** vorhanden, aber die Variablen können auch Zeilen eingerichtet werden. Ein Datensatz kann aus mehreren Zellblöcken bestehen, wodurch Sie dann Daten in verschiedene Blätter derselben Arbeitsmappe eingeben können.

Wenn Sie einen Datensatz definieren, versucht StatTools, in Excel die Variablen zu identifizieren, die sich in den Zellen um die aktuelle Auswahl befinden. Dadurch kann schnell und mühelos ein Datensatz eingerichtet werden, und zwar mit den Variablen in Spalten und den entsprechenden Namen der Variablen jeweils in der obersten Zeile.

In Ihrem Datensatz brauchen die Spalten nicht unbedingt alle gleich lang zu sein. Sie könnten beispielsweise zwei Variablen – *Gewicht_Männer* und *Gewicht_Frauen* – mit unterschiedlicher Anzahl von Beobachtungen haben. Bei vielen Analysen wird StatTools jedoch die leeren Zellen in den kürzeren Spalten einfach als fehlende Daten behandeln.

Gestapelte und entstapelte Daten

StatTools unterstützt sowohl gestapelte als auch entstapelte Daten. Bei einigen statistischen Vorgängen ist es einfacher, mit gestapelten Daten zu arbeiten. Bei anderen Vorgängen sind dagegen evtl. entstapelte Daten vorteilhafter. Wenn wir beispielsweise die durchschnittlichen Haushaltseinkommen in mehreren unterschiedlichen Nachtbarschaften vergleichen, würde bei der entstapelten Form eine separate *Einkommens*-Variable (oder Spalte) für jede Nachtbarschaft vorhanden sein. Bei der gestapelten Form würde dagegen die Wertevariable *Einkommen* und die Kategorievariable *Nachbarschaft* verfügbar sein, um zu erkennen zu geben, in welcher Nachbarschaft sich die einzelnen Haushalte befinden.

Die Variablen-Stapelfunktion in StatTools ermöglicht Ihnen, Ihre Variablen in zwei Spalten zu stapeln, und zwar in der Wertespalte *Einkommen* und der Kategoriespalte *Nachbarschaft*. Je nach Analysentyp könnte es vielleicht einfacher sein, mit dem gestapelten Datensatz zu arbeiten, als mit der entstapelten Version.

Behandlung von fehlenden Werten

Wenn in Ihrem Datensatz fehlende Werte vorhanden sind (was bei der Statistikanalyse häufig vorkommt), werden diese der Aufgabe entsprechend gehandhabt. Bei Übersichtsfaktoren, wie z.B. Mittelwerten und Standardabweichungen, werden fehlende Werte beispielsweise einfach ignoriert. Bei einer Regressionsanalyse mit drei Variablen werden dagegen im Datensatz für diese drei Variablen jeweils nur solche Zeilen verwendet, die keine fehlenden Werte enthalten. (Dies wird „listenweise" oder „fallweise" Löschung genannt.) Als drittes Beispiel kann ein Punktdiagramm mit zwei Variablen genannt werden, in dem nur solche Punkte grafisch dargestellt werden, bei denen beide Variablen Werte haben, die nicht zu den fehlenden gehören.

Hinweis: Bei einigen StatTools-Vorgängen kann nicht mit fehlenden Werten gearbeitet werden. Im Abschnitt* Referenz *dieses Handbuchs wird beschrieben, wie bei den einzelnen Vorgängen bei fehlenden Werten vorgegangen wird.

Berichte und Diagramme in StatTools

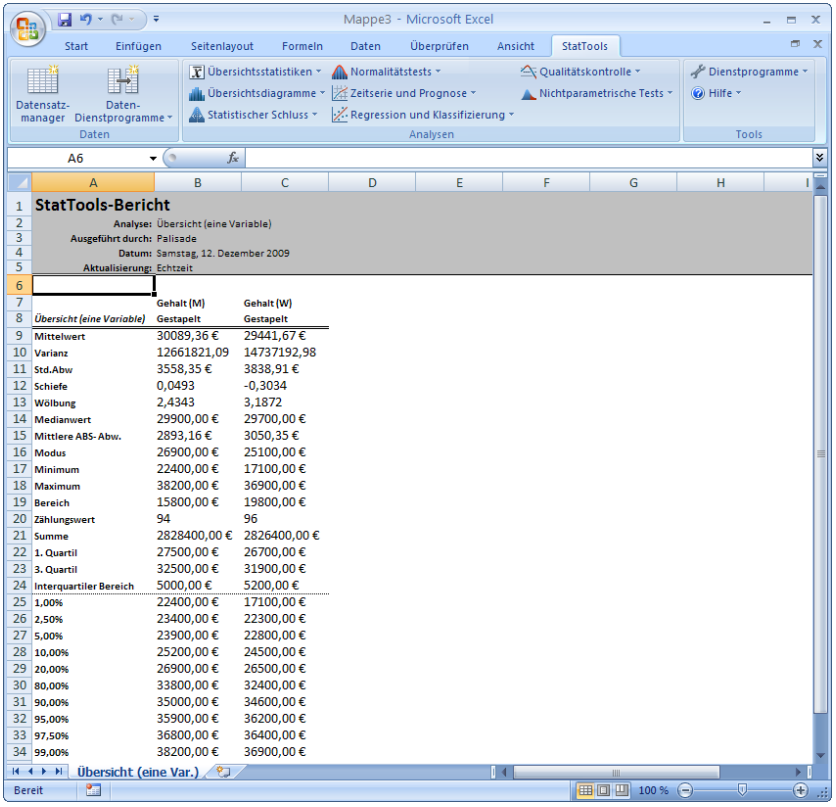
Wenn in StatTools eine numerische Ausgabe erstellt wird, wie z.B. ein Bericht aus einer Regressionsanalyse oder eine Übersichtstatistikstabelle, werden folgende Optionen zur Platzierung des Berichts gegeben:

- **In neuer Arbeitsmappe –** was bedeutet, dass nötigenfalls eine neue Arbeitsmappe erstellt und dann jeder Bericht darin auf einem neuen Arbeitsblatt platzieret wird.
- **Auf neuem Arbeitsblatt in aktiver Arbeitsmappe –** was bedeutet, dass jeder Bericht in der aktiven Arbeitsmappe auf einem neuen Arbeitsblatt platziert wird.
- **Beginn nach zuletzt verwendeter Spalte –** was bedeutet, dass jeder Bericht rechts der zuletzt verwendeten Spalte auf einem aktiven Arbeitsblatt platziert wird.
- **Beginn in Zelle –** was bedeutet, dass Sie eine Zelle auswählen können, in der dann die linke obere Kante des Berichts oder Diagramms platziert wird.

Wenn durch StatTools irgendwelche Diagramme erstellt werden, sind diese bei den Berichten zu sehen. Diagramme werden im Excel-Format erstellt und können über die standardmäßigen Excel-Diagrammbefehle Ihren Wünschen angepasst werden.

Verwendung von Formeln gegenüber Werten

StatTools versucht in der Regel, die Ergebnisse möglichst in Echtzeit anzuzeigen. Mit anderen Worten, Berichte werden möglichst durch Formeln mit den Originaldaten verknüpft. Angenommen, Sie arbeiten mit der Variable *Gewicht* und möchten Übersichtsfaktoren über das *Gewicht*, wie z.B. Mittelwert und Standardabweichung, haben. Durch den Übersichtsstatistikvorgang erhält der Gewichtsbereich den Namen *Gewicht* und werden dann folgende Formeln in die Ausgabezellen eingegeben: **=StatMean(Gewicht)** und **=StatStdDev(Gewicht)**. Bei StatMean und StatStdDev handelt es sich um in StatTools integrierte Funktionen, die zum Berechnen des Mittelwerts und der Standardabweichung verwendet werden. Dadurch werden die standardmäßigen für diese Statistiken bestimmten Excel-Funktionen ersetzt.



StatTools-Bericht

Analyse:	Übersicht (eine Variable)
Ausgeführt durch:	Palisade
Datum:	Samstag, 12. Dezember 2009
Aktualisierung:	Echtzeit

Übersicht (eine Variable)	Gehalt (M) Gestapelt	Gehalt (W) Gestapelt
Mittelwert	30089,36 €	29441,67 €
Varianz	12661821,09	14737192,98
Std.Abw	3558,35 €	3838,91 €
Schiefe	0,0493	-0,3034
Wölbung	2,4343	3,1872
Medianwert	29900,00 €	29700,00 €
Mittlere ABS- Abw.	2893,16 €	3050,35 €
Modus	26900,00 €	25100,00 €
Minimum	22400,00 €	17100,00 €
Maximum	38200,00 €	36900,00 €
Bereich	15800,00 €	19800,00 €
Zählungswert	94	96
Summe	2828400,00 €	2826400,00 €
1. Quartil	27500,00 €	26700,00 €
3. Quartil	32500,00 €	31900,00 €
Interquartiler Bereich	5000,00 €	5200,00 €
1,00%	22400,00 €	17100,00 €
2,50%	23400,00 €	22300,00 €
5,00%	23900,00 €	22800,00 €
10,00%	25200,00 €	24500,00 €
20,00%	26900,00 €	26500,00 €
80,00%	33800,00 €	32400,00 €
90,00%	35000,00 €	34600,00 €
95,00%	35900,00 €	36200,00 €
97,50%	36800,00 €	36400,00 €
99,00%	38200,00 €	36900,00 €

Es gibt zwei Gründe dafür, dass Formeln in Berichten verwendet werden. Erstens werden Sie dadurch mit den statistischen Vorgängen und den StatTools-Funktionen in Excel vertraut, da Sie nicht nur die numerischen Ergebnisse sehen, sondern auch erkennen können, wie diese formuliert werden. Und zweitens ist dadurch auch der praktische Vorteil vorhanden, dass sich die Ergebnisse bei Datenänderung automatisch aktualisieren, sodass Sie den Vorgang nicht noch einmal erneut ausführen brauchen.

Mitunter ist es jedoch nicht angebracht, Formeln zu verwenden. Ein gutes Beispiel dafür ist die Regression. StatTools gibt Ihnen nicht die Formeln, die zum Erstellen der Regressionsausgabe verwendet werden, sondern nur die numerischen Ergebnisse. In solchen Fällen müssen Sie also bei Datenänderung die Vorgänge erneut ausführen.

StatTools gibt Ihnen auch die Möglichkeit, die Echtzeit-Aktualisierung auszuschalten. Das ist praktisch, wenn in Excel die Neuberechnung bei Datenänderung zu lange dauert.

Verwendung von Kommentaren in Zellen

In Excel haben Sie die Möglichkeit, in Zellen so genannte Popup-Kommentare zu verwenden. Wenn sich ein Kommentar in einer Zelle befindet, ist das durch ein kleines rotes Dreieck oben rechts in der Zelle zu erkennen. Um den Kommentar lesen zu können, müssen Sie langsam mit dem Cursor über die Zelle gehen. In StatTools werden diese Kommentare für kontextbezogene Hilfe verwendet. Dies ist sozusagen die am schnellsten verfügbare Online-Hilfe.

Übrigens, falls die Kommentare in einer Kalkulationstabelle stets angezeigt bleiben, d.h., ständig vor den Daten zu sehen sind, sollten Sie im Menü **Extras** den Befehl **Optionen** wählen, dann auf die Registerkarte **Ansicht** und anschließend auf die Schaltfläche **Nur Kommentarindikator** klicken.

Kapitel 3: StatTools-Referenzhandbuch

Einführung **23**

VBA-Makrosprache und Toolkit für Entwickler in StatTools **23**

Referenz: StatTools-Symbole **25**

StatTools-Symbolleiste **25**

Referenz: StatTools-Menübefehle **27**

Einführung **27**

Auflistung der Befehle **28**

StatTools-Menü – Datensätze **95**

Befehl *Datensatzmanager* **95**

Menü *Dienstprogramme* **101**

Befehl *Stapeln* **101**
Befehl *Entstapeln* **104**
Befehl *Daten transformieren* **105**
Befehl *Verzögerung* **107**
Befehl *Differenz* **109**
Befehl *Interaktion* **111**
Befehl *Kombination* **113**
Befehl *Pseudo* **115**
Befehl *Zufallsprobe* **117**

Menü *Übersichtsstatistik* **119**

Befehl *Übersicht (eine Variable)* **119**
Befehl *Korrelationen und Kovarianz* **122**

Menü *Übersichtsdiagramme* **125**

Befehl *Histogramm* **125**
Befehl *Punktdiagramm* **128**
Befehl *Box-Whisker-Plot* **131**

Menü *Statistischer Schluss* **135**

Befehl *Vertrauensintervall – Mittelwert/Std. Abweichung* **135**
Befehl *Vertrauensintervall für Proportionen* **138**
Befehl *Hypothesetest für Mittelwert/Std. Abweichung* **141**
Befehl *Hypothesetest für Proportion* **144**
Befehl *Probegrößenauswahl* **148**
Befehl *Einseitige ANOVA* **150**
Befehl *Zweiseitige ANOVA* **154**
Befehl *Chi-Quadrat-Unabhängigkeitstest* **157**

Menü *Normalitätstests* **161**

Befehl *Chi-Quadrat-Normungstest* **161**
Befehl *Lilliefors-Test* **165**
Befehl *Normales Q-Q-Diagramm* **167**

Menü *Zeitserie und Prognose* **169**

Befehl *Zeitserien-Diagramm* **169**
Befehl *Autokorrelation* **172**
Befehl *Ausführzählungstest für Zufallsfaktoren* **174**
Befehl *Prognose* **176**

Menü *Regression und Klassifizierung* **181**

Befehl Regression **181**
Befehl *Logistische Regression* **193**
Befehl *Diskriminanzanalyse* **198**

Menü *Qualitätskontrolle* **203**

Befehl *Pareto-Diagramm* **204**
Befehl *x/r-Diagramme* **207**
Befehl *p-Diagramm* **211**
Befehl *c-Diagramm* **215**
Befehl *u-Diagramm* **218**

Menü *Nichtparametrische Tests* **223**

Befehl *Vorzeichentest* **225**
Befehl *Wilcoxon-Vorzeichen-Rangtest* **228**
Befehl *Mann-Whitney-Test* **231**
Befehl Kruskal-Wallis-Test **235**

Menü *multivariate Analyse* **239**

Befehl Hauptkomponenten-Analyse **239**
Befehl Clusteranalyse **243**

Menü *Dienstprogramme* **251**

Befehl *Anwendungseinstellungen* **251**
Befehl *Datensätze löschen* **258**

Befehl *Dialogfeldspeicher löschen* **258**
Befehl *StatTools-Add-In entladen* **258**

Hilfemenü **259**

StatTools-Hilfe **259**
Befehl *Online-Handbuch* **259**
Befehl *Lizenzaktivierung* **259**
Befehl *Info über* **259**

Referenz: StatTools-Funktionen **261**

Einführung **261**

StatTools-Funktionen gegenüber Excel-Funktionen **261**
Verteilungsfunktionen **263**
Echtzeit-Berichte **264**

Referenz: Liste der Statistikfunktionen **267**

Tabelle der verfügbaren Funktionen **267**
Detaillierte Funktionsbeschreibungen **270**

Einführung

In diesem Kapitel werden die in StatTools verwendeten Symbole, Befehle und Statistikfunktionen beschrieben. Das Kapitel ist in vier Abschnitte unterteilt:

1) *Referenz: StatTools-Symbole*
2) *Referenz: Übersicht über StatTools-Vorgänge*
3) *Referenz: StatTools-Befehle*
4) *Referenz: StatTools-Funktionen*

VBA-Makrosprache und Toolkit für Entwickler in StatTools

StatTools enthält eine leistungsstarke VBA-basierte Makrosprache, die dazu verwendet werden kann,

1) **die StatTools-Analysen zu automatisieren**
2) **neue Statistikanalysen zu entwickeln, in denen StatTools-Datensatzmanager, -Berichte und -Diagramme verwendet werden** Durch diese benutzerdefinierten Berechnungen können Analysen erstellt werden, die in den eigentlichen StatTools-Vorgängen nicht vorhanden sind. Falls erwünscht, können diese Analysen auch im StatTools-Menü und den StatTools-Symbolleisten angezeigt werden.

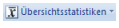
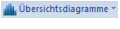
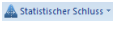
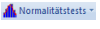
Weitere Informationen über die StatTools VBA-Makrosprache und das Toolkit für Entwickler sind in der Online-Dokumentation zu finden, die für das Produkt verfügbar ist.

Referenz: StatTools-Symbole

StatTools-Symbolleiste

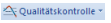
StatTools-Symbole werden dazu verwendet, Datensätze und Variablen zu definieren und diese Variablen dann statistisch auszuwerten. StatTools-Symbole sind in der Excel-Symbolleiste bzw. in einer Befehlsleiste in Excel 2007 zu finden. In diesem Abschnitt wird jedes Symbol kurz beschrieben und werden dabei die Funktionen, die dadurch ausgeführt werden, dargelegt. Auch werden die entsprechenden Menübefehle beschrieben.

In Excel 2007 sind in der StatTools-Befehlsleiste folgende Symbole zu sehen.

Symbol	*Symbolfunktion und entsprechender Befehl*
	Datensatz und Variablen definieren oder einen bestehenden Datensatz nebst Variablen bearbeiten bzw. löschen *Entsprechender Befehl: Datensatzmanager*
	Dienstprogramm (für fehlende Daten) ausführen *Entsprechender Befehl: Dienstprogramm für fehlende Daten*
	**Übersichtsstatistik ausführen** *Entsprechender Befehl: Übersichtsstatistiken*
	**Übersichtsdiagramme für Variablen erstellen** *Entsprechender Befehl: Übersichtsdiagramme*
	**Statistischen Schluss ausführen** *Entsprechender Befehl: Statistischer Schluss*
	**Normalitätstest an Variablen ausführen** *Entsprechender Befehl: Normalitätstest*
	Zeitserie oder Prognose ausführen *Entsprechender Befehl: Zeitserie und Prognose*

Regression oder Klassifizierung ausführen

Entsprechender Befehl: Regression und Klassifizierung

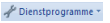


Qualitätskontrolle ausführen

Entsprechender Befehl: Qualitätskontrolle

Nichtparametrischen Test ausführen

Entsprechender Befehl: Nichtparametrische Tests



StatTools-Dienstprogramme anzeigen

Entsprechender Befehl: Dienstprogramme

StatTools-Hilfedatei anzeigen

Entsprechender Befehl: Hilfe

Referenz: StatTools-Menübefehle

Einführung

Dieser Abschnitt des Referenzhandbuchs beschäftigt sich im Einzelnen mit den verfügbaren StatTools-Befehlen, in Excel 2007 bzw. späteren Excel-Versionen in der StatTools-Befehlsleiste zu sehen sind. Die Befehle werden in der Reihenfolge erörtert, in der sie in der Menüleiste erscheinen, d. h. angefangen mit dem Befehl Datensatzmanager und von da aus nach rechts gehend. Viele der verfügbaren Befehle können auch über die StatTools-Symbole ausgeführt werden. Im Abschnitt **Referenz: StatTools-Symbole** in diesem Kapitel sind die Befehle aufgeführt, die den einzelnen StatTools-Symbolen entsprechen.

Mehrere StatTools-Befehle sind auch in dem unverankerten Kontextmenü verfügbar, das erscheint, wenn Sie in Excel mit der rechten Maustaste klicken.

Auflistung der Befehle

Die in StatTools verfügbaren Vorgänge sind in naturgemäße Gruppen unterteilt. Für jede Gruppe befindet sich ein Menüeintrag im StatTools–Menü. Falls eine Gruppe mehr als einen Eintrag hat, ist ein Untermenü vorhanden, in dem die Einträge aufgelistet sind. In diesem Abschnitt wird jeder Vorgang in den einzelnen Gruppen kurz beschrieben. Detailliertere Informationen über jeden Vorgang sind in diesem Kapitel unter **Referenz: StatTools-Befehle** zu finden.

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenan-forderungen	Daten aus mehreren Zellbe-reichen?	Ungül-tige Daten	Anzahl der Var.
Übersichtsstatistiken							
Befehl Übersicht (eine Variable)	Erstellt Übersichtsstatistiken einschließlich der üblichen Messwerte, wie z.B. Durchschnitt, Medianwert und Standardab-weichung, und der Optionen, wie z.B. Quartile und Perzentile.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-100
Befehl Korrelationen und Kovarianz	Erstellt eine Korrelations- oder Kovarianztabelle für von Ihnen ausgewählte Variablen.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-250
Übersichtsdiagramme							
Befehl Histogramm	Erstellt ein Histogramm für jede von Ihnen ausgewählte Variable. Gibt Ihnen die Möglichkeit, die Kategorien oder Bins des Histogramms zu definieren.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Teilweise in Echtzeit – Diagramm wird durch Datenänderungen aktualisiert, wenn sich die Daten innerhalb des x-Achsen-bereichs des Diagramms befinden	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-100

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenan-forderungen	Daten aus mehreren Zellbe-reichen?	Ungül-tige Daten	Anzahl der Var.
Befehl Punktdiagramm	Erstellt ein Punktdiagramm für jedes von Ihnen ausgewählte Variablenpaar.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Nur entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1-10
Befehl Box-Whisker-Plot	Erstellt ein Box-Plot (wenn Sie nur eine Variable auswählen) oder nebeneinander befindliche Box-Plots (wenn Sie mehrere Variablen auswählen).	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Nein	Ignoriert	1-10
Statistischer Schluss							
Befehl Vertrauensintervall – Mittelwert/Std. Abweichung	Berechnet bei einzelnen Variablen ein Vertrauens-intervall für Mittelwert und Standardabweichung oder die Differenz zwischen den Mittelwerten, wenn es sich um Variablenpaare handelt. Die Vertrauensintervalle können mithilfe einer Analyse berechnet werden, und zwar unter Verwendung einer, zweier oder gepaarter Werteproben.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)
Befehl Vertrauensintervall für Proportionen	Analysiert die Proportion von Elementen in einer Werteprobe, die zu einer bestimmten Kategorie gehört (Analyse unter Verwendung nur einer Werteprobe) oder vergleicht zwei Werteproben hinsichtlich der Proportion von Elementen in einer bestimmten Kategorie (Analyse unter Verwendung von zwei Werteproben).	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellbereichen?	Ungültige Daten	Anzahl der Var.
Befehl Hypothesetest für Mittelwert/Std. Abweichung	Führt bei einzelnen Variablen einen Hypothesetest für Mittelwert und Standardabweichung oder für die Differenzen zwischen den Mittelwerten aus, wenn es sich um Variablenpaare handelt. Die Hypotheseteste können mithilfe einer Analyse ausgeführt werden, und zwar unter Verwendung von einer, zwei oder gepaarten Werteproben.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)
Befehl Hypothesetest für Proportion	Analysiert die Proportion von Elementen in einer Werteprobe, die zu einer bestimmten Kategorie gehört (Analyse unter Verwendung nur einer Werteprobe) oder vergleicht zwei Werteproben hinsichtlich der Proportion von Elementen in einer bestimmten Kategorie (Analyse unter Verwendung von zwei Werteproben).	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)
Befehl Probegrößenauswahl	Bestimmt die Werteprobengröße (oder Probengrößen) die erforderlich sind, um ein Vertrauensintervall mit einer vorgeschriebenen Halblänge zu erhalten. Dadurch werden die Vertrauensintervalle für Mittelwert, Proportion, Differenz zwischen Mittelwerten und Differenz zwischen zwei Proportionen festgelegt.	n/b	n/b	n/b	n/b	n/b	n/b
Befehl Einseitige ANOVA	Eine Erweiterung einer Analyse unter Verwendung von zwei Werteproben, um die Mittelwerte von zwei Populationen zu vergleichen. Dadurch wird getestet, ob mehrere Mittelwerte gleich sind.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	2-50

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellbereichen?	Ungültige Daten	Anzahl der Var.
Befehl Zweiseitige ANOVA	Führt eine zweiseitige Varianzanalyse aus. Diese Analyse wird gewöhnlich im Zusammenhang mit einer experimentellen Planung ausgeführt, die zwei Faktoren enthält, die beide auf mehrere „Behandlungsstufen“ eingestellt sind.	Nicht zulässig	Echtzeit	Gestapelte Daten Bis zu 16 Millionen Fälle möglich Experiment muss ausgeglichen sein	Ja	Nicht zulässig	2 Kategorie-variablen, 1 Wertvariable
Befehl Chi-Quadrat-Unabhängigkeitstest	Verwendet einen Chi-Quadrat-Test, um zu prüfen, ob die Zeilen- und Spaltenattribute in einer Eventualfalltabelle statistisch unabhängig sind.	Nein	Echtzeit (solange sich die Tabellengröße nicht ändert)	n/b	n/b	n/b	n/b
Normalitätstests							
Befehl Chi-Quadrat-Normungstest	Führt einen Chi-Quadrat-Normungstest für die von Ihnen gewählte Variable aus.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Teilweise in Echtzeit (Bin-Position bleibt unverändert, aber Belegung und Diagramme ändern sich)	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1
Befehl Lilliefors-Test	Liefert einen leistungsstärkeren Normalitätstest als durch den Chi-Quadrat-Anpassungsgütetest möglich ist. („Leistungsstärker“ bedeutet in diesem Fall, dass durch den Test besser irgendeine nicht normale Anpassung erkannt werden kann.)	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-10
Befehl Normales Q-Q-Diagramm	Erstellt ein Q-Q-Diagramm für eine ausgewählte Variable und sorgt für einen informellen Normalitätstest.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Nein	Ignoriert	1
Zeitserie und Prognose							
Befehl Zeitserien-Diagramm	Erstellt ein Zeitseriendiagramm für eine oder mehrere Zeitserien-Variablen, die sich alle in derselben Tabelle befinden.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1-100

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellbereichen?	Ungültige Daten	Anzahl der Var.
Befehl Autokorrelation	Berechnet eine beliebige Anzahl von Autokorrelationen für eine Zeitserien-Variable, gibt an, welche davon ungleich Null sind und (falls gewünscht) zeigt ein Balkendiagramm (d.h., ein so genanntes „Korrelogramm“) der Autokorrelationen an.	Zu Beginn und Ende der Daten zulässig	Echtzeit	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1-10
Befehl Ausführzählungstest für Zufallsfaktoren	Führt einen Ausführzählungstest aus, um festzustellen, ob eine Variable (gewöhnlich eine Zeitserien-Variable) eine Zufallsvariable ist.	Zu Beginn und Ende der Daten zulässig	Echtzeit	Entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Nicht zulässig	1 oder mehrere
Befehl Prognose	Sagt die Zeitseriendaten voraus, und zwar mittels gleitender Durchschnittsmethode, einfachem exponentiellem Glätten, exponentieller Glättungsmethode für erfasste Tendenz (Holt) und exponentieller Glättungsmethode für erfasste Saisonabhängigkeit (Winters).	Nur zu Datenbeginn zulässig	Echtzeit	Entstapelte Daten Bis zu 32.000 Fälle möglich	Ja	Nicht zulässig	1 oder mehrere
Regression und Klassifizierung							
Error! Reference source not found.	Führt eine Reihe von Regressionsanalysen aus, einschließlich einfacher Mehrfachregression, schrittweiser Regression, Vorwärtsregression, Rückwärtsregression und Blockregression.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Statisch	Unstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Nicht zulässig	1 abhängige, 1 – 250 unabhängige
Befehl Regression Führt Regressionsanalysen an einem Satz von Variablen aus Die Regressionsanalyse unterstützt verschiedene	Führt eine logistische Regressionsanalyse an einem Datensatz aus. Dies ist im Wesentlichen eine nicht lineare Art von Regressionsanalyse, wobei dann die Reaktionsvariable binär ist, d.h., 0 oder 1.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Statisch	Entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Nicht zulässig	1 abhängige, 1 – 250 unabhängige

Regressionsmodelle, einschließlich Pseudovariablen-Regression (Regression mit kategorischen Variablen), Polynom-Regression und Regression mit Interaktionen zwischen Variablen. Mit dem optionalen Regressionsassistenten können Sie diese Modelle definieren, falls Sie diese verwenden möchten. Für jedes dieser Modelle sind verschiedene Methoden zur Auswahl der Variablen verfügbar, die in die							

Gleichung aufgenommen werden sollen. Diese werden als „Regressionstypen“ bezeichnet und dabei handelt es sich um die Methoden Mehrfache Regression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression und Blockregression. Berichte über die einzelnen Regressionen enthalten folgende Einzelheiten: eine ANOVA-Tabelle und eine Tabelle der geschätzten Regressionskoeffizienten nebst zugehörigem Standardfehler, t-Werten, p-Werten und							

auch entsprechende 95-prozentige Vertrauensbereiche. Auch haben Sie die Möglichkeit, zwei neue Variablen zu erstellen, d. h., für angepasste und Residualwerte, sowie eine Anzahl von diagnostischen Punktdiagrammen und anderen weiter unten beschriebenen Optionen zu erstellen. ***Regressions-assistent*** Der Regressionsassistent bietet eine einfache Möglichkeit zum Erstellen von Regressionsmodellen wie Pseudovaria							

blen-Regression oder Polynom-Regression. Für diese Modelle müssen Variablen-Transformationen in der Analyse miteingeschlossen werden („abgeleitete Variablen“); der Assistent erleichtert das Miteinschließen der Transformationen. Der Assistent führt Sie durch vier Schritte: **Schritt 1: Modellauswahl** – Hier geben Sie den Typ des Modells an, das Sie erstellen möchten. Es sind folgende Optionen verfügbar:							

• **Standardregressionsmodell** – Mehrfachregressionsmodell. • **Pseudovariablen-Regressionsmodell** – Mehrfachregressionsmodell, das kategoriell unabhängige Variablen enth							

ält. Die 0/1-Codierung für die Kategorien wird automatisch durchgeführt.

- **Polynom-Regressionsmodell** – Mehrfachregressionsmodell, das die Güte der unabhängig

en Variablen enthält. Es können auch alle möglichen Interaktionen zwischen den ursprünglichen Variablen berücksichtigt werden. • **Regressionsmodell mit Inte**							

raktionen zwischen Variablen – Mehrfachregressionsmodell, das alle möglichen Interaktionen zwischen unabhängigen Variablen enthält. • **Regressionsmodell mit nic**							

htli nea ren Var iabl entr ansf orm atio nen – Reg ress ions mo dell, bei dem eini ge nich tline are Tra nsfo rma tion en auf die Vari able n ang ewa ndt wer den. Es sind folg end e Opt ione							

n verfügbar: **Schritt 2**: Auswahl von unabhängigen Variablen – Hier wählen Sie eine oder mehrere unabhängige Variablen und etwaige erwünschte Variablentransformationen aus. **Schritt 3**: **Auswahl der abhängigen Variable –** Hier wählen Sie die abhängige Variable aus. Bei einigen Modelltype							

n wird auch eine Transformation der abhängigen Variable ausgewählt. **Schritt 4**: Auswahl der Auswahlmethode – Hier wählen Sie den „Regressionstyp“ aus – Mehrfachregression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression oder Blockregression. Wenn der Assistent fertig ist, können die Einstellungen im Dialogfeld Regression verändert werden. ***Verfügbare Regressions-typen (Methoden der Variablenauswahl)***							

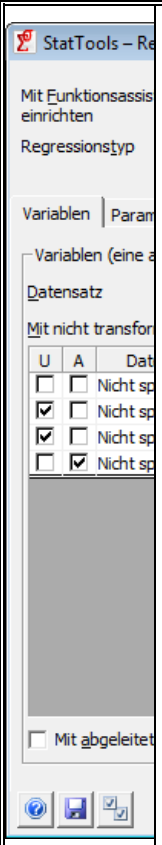
Bei den verfügbaren Regressionstypen (Methoden zur Variablenauswahl) handelt es sich um **Mehrfachregression**, **Schrittweise Regression**, **Vorwärtsregression**, **Rückwärtsregression** und **Blockregression**. Durch die Mehrfachregression wird sofort eine komplette Gleichung erstellt, und zwar unter Verwendung der ausgewählten unabhängigen Variablen. (Das umfasst den Vorgang, der gewöhnlich als							

„einfache“ Regression bezeichnet wird, bei dem es eine einzige unabhängige Variable gibt.) Bei den anderen Regressionsverfahren können die Variablen (oder Variablenblöcke) sequenziell der Gleichung hinzugefügt oder aus der Gleichung entfernt werden. Insbesondere können Variablen bei der schrittweisen Regression eine nach der anderen hinzugefügt werden. Dabei ist die als nächste hinzuzufügende immer die Variable, die am stärksten mit dem							

unerklärten Teil der abhängigen Variable korreliert ist. Bei der schrittweisen Option können Variablen nach der Hinzufügung außerdem auch wieder entfernt werden, wenn sie keinen erheblichen Beitrag mehr leisten können. Die Vorwärtsregression entspricht praktisch der schrittweisen Regression, mit dem Unterschied, dass die Variablen bei der Vorwärtsregression nach Hinzufügung nicht wieder entfernt werden können. Bei							

der Rückwärtsregression sind zu Beginn alle möglichen unabhängigen Variablen in der Gleichung enthalten und werden dann einzeln gelöscht, sobald sie keinen bedeutsamen Beitrag mehr leisten. Die Blockregression ermöglicht es schließlich, dass unabhängige Variablen in Form von *Blöcken* hinzugefügt oder nicht hinzugefügt werden, und zwar in einer bestimmten Reihenfolge. Wenn ein Block nicht signifikant							

ist und deshalb nicht hinzugefügt wird, werden auch keine späteren Blöcke mehr berücksichtigt. Zum Definieren dieser Analysen wird das Dialogfeld **Regression** verwendet: ***Dialogfeld Regression***							



Im Dialogfeld Regression stehen folgende Optionen zur Verfügung:

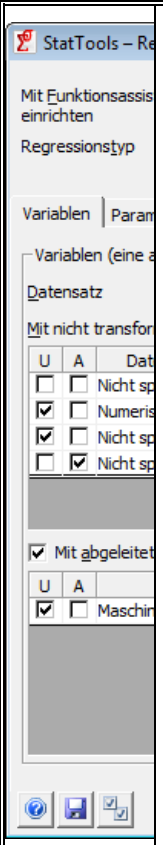
- **Mit Funktionsassistent einrichten –** Mit dies

er Option werden Sie in einer Folge von Dialogfeldern durch den Vorgang zum Erstellen verschiedener Regressionsmodelle geführt. • **Regressionstyp** – Hie							

r wählen Sie die Methode zur Variablenauswahl aus: Mehrfache regression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression oder Blockregression.							

• **Unabhängige Variablen festlegen –** Wenn Sie diese Option aktivieren, enthält der Variablen-Auswahlraster eine zusätzliche Spalte F. Die Spalte F ermöglicht Ihnen							

die Auswahl „fixer“ Variablen. Das sind Variablen, die beim Variablen-Auswahlvorgang niemals aus der Gleichung ausgeschlossen werden.							
Dialogfeld Regression – Registerkarte Variablen							

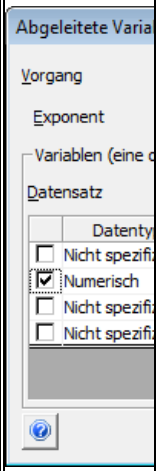
 Für diese Analysen müssen eine abhängige Variable (D) und eine oder mehrere unabhängige Variablen (I) ausgewählt werden, aber nicht im Fall der Blockregression. Bei der Blockregression ist es							

erforderlich, eine abhängige Variable (D) und ein bis sieben Blöcke (B1 bis B7) unabhängiger Variablen auszuwählen. Da kategorische Variablen transformiert werden müssen, bevor sie in einer Regressionsanalyse verwendet werden können, müssen ausgewählte Variablen entweder als „kategorische“ oder „numerische“" Variablen bezeichnet werden. Bei dem ausgewählten Datensatz muss es sich immer um einen entstapelten							

handeln. Die Variablen können dabei aus verschieden en Datensätzen stammen. Die Option für abgeleitete Variablen ist jedoch bei mehreren Datensätzen nicht verfügbar. Wenn das Kontrollkäst chen **Mit abgeleitete n Variablen** aktiviert ist, können Transformat ionen von Variablen – wie das Quadrat einer Variable – der Analyse hinzugefügt werden. Wenn Sie auf die Schaltfläche **Hinzufügen** klicken, wird das Dialogfeld **Abgeleitete**							

Variablen hinzufügen angezeigt.

Dialogfeld Abgeleitete Variablen hinzufügen



Das Dialogfeld **Abgeleitete Variablen hinzufügen** bietet eine schnelle und einfache Methode zum Erstellen neuer Variablen, die Transformationen der ursprünglichen Variablen sind. Es sind folgende

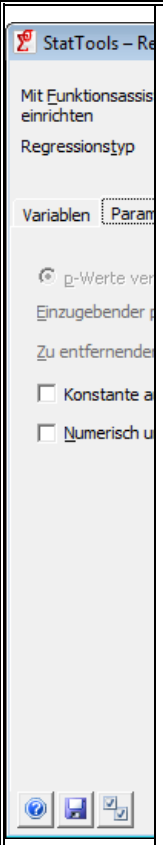
Optionen verfügbar: • **Güte –** Wählen Sie diese Option aus, um Gütetransformationen auf die Variablen anzuwenden. • **Logarithmus –** Wählen Sie diese Option aus, um logarithmis							

che Transformationen auf die Variablen anzuwenden.

- **Interaktion ohne Kategorievariable –** Wählen Sie diese Option aus, um einen Interaktionsbegriff hinzuzu

fügen, der das Produkt von zwei oder mehr numerischen Variablen ist. • **Interaktion mit Kategorievariable –** Wählen Sie diese Option aus, um einen Inte							

raktionsbegriff auf Basis von einer oder mehreren numerischen Variablen und einer kategorischen Variable hinzuzufügen.							
Dialogfeld Regression – Registerkarte Parameter							



Für Parameter sind folgende Optionen verfügbar:

- **p-Werte verwenden** – Wirkt sich darauf aus,

wie die Variablen der Regressionsgleichung hinzugefügt oder daraus entfernt werden. Wenn diese Option ausgewählt ist, können Sie jeden verwendeten Regress							

ionsmethode angeben, dass ein p-Wert eingegeben oder entfernt werden soll. Je geringer der p-Wert ist, desto wichtiger muss eine Variable sein, um zur Regress							

ions glei chu ng hinz ugef ügt ode r dar aus entf ernt wer den zu kön nen. Die gez eigt en Stan dar dwe rte sind gew öhn lich akz epta bel. Sie müs sen jedo ch zwe i Din ge ber ück sich							

tigen. 1) Der einzugebende p-Wert darf nicht größer als der zu entfernende p-Wert sein. 2) Um Variablen leichter hinzufügen (und nicht so leicht entfernen) zu							

können, sollten größere p-Werte verwendet werden. Um Variablen dagegen schwieriger hinzufügen (und leichter entfernen) zu können, sollten kleinere p-Werte							

verwendet werden. p-Werte liegen gewöhnlich im Bereich von 0,01 bis 0,1.

- **f-Werte verwenden** – Wirkt sich (wie p-Werte) darauf aus, wie die Variablen der Reg

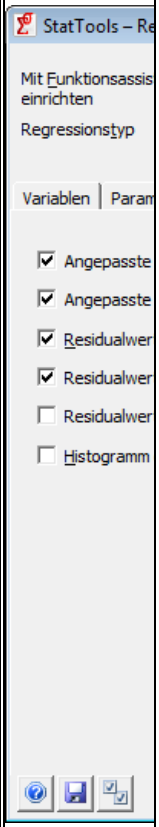
ressionsgleichung hinzugefügt oder daraus entfernt werden. Diese Option ermöglicht Ihnen, je nach verwendeter Regressionsmethode anzugeben, dass ein f-Wert							

eingegeben oder entfernt werden soll. f-Werte liegen gewöhnlich im Bereich von 2,5 bis 4.

- **Konstante auf Null einstellen –** Erzwingt, dass die Regression durch

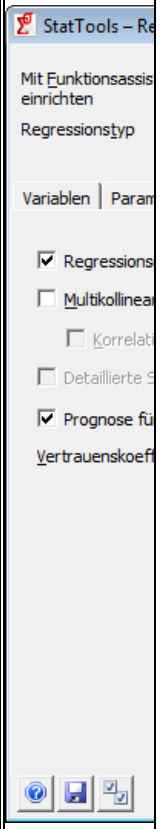
den Ausgangspunkt verläuft. • **Numerisch unabhängige Variablen standardisieren –** Der Mittelwert wird von jeder Variable abgezogen und das Ergebnis dan							

n durch die Standardabweichung dividiert. Das ist manchmal von Vorteil, wenn die unabhängigen Variablen sehr unterschiedliche Größenordnungen haben.							

Dialogfeld Regression – Registerkarte Diagramme  Durch StatTools können mehrere optionale Punktdiagramme erstellt werden, wie im Dialogfeld unter Diagramme zu sehen ist. Dabei							

handelt es sich u. a. um: • **Angepasste Werte gegenüber aktuellen y-Werten** • **Angepasste Werte gegenüber x-Werten** • **Residualwerte gegenüber angepassten Wer**							

ten • **Residualwerte gegenüber x-Werten** • **Residualwerte gegenüber Ordnungs-Index** • **Histogramm der Residualwerte** Diese Diagramme werden in der Regel bei der Residualwert-Analyse verwendet,							

um zu überprüfen, ob den Regressionsannahmen entsprochen wird. ***Dialogfeld Regression – Registerkarte Optionen***  Unter den erweiterten Optionen für die Regressionsanalyse ist Folgendes zu finden:							

• **Regressionsgleichung anzeigen –** Dies verdeutlicht die Regressionsgleichungen für den Bericht. Es ist nützlich, wenn eine oder mehrere kategorische Variablen ber							

ücksichtigt werden. In diesem Fall wird für jede Kombination der Kategorien eine eigene Gleichung gezeigt. • **Multikollinearität prüfen** – Das umfasst den Varianz-Hoc							

hrechnungsfaktor (VIF) der einzelnen unabhängigen Variablen im Bericht. Große VIF-Werte weisen auf Multikollinearität hin. Die Korrelations-Matrix kann auc							

h berücksichtigt werden. • **Detaillierte Schrittinformationen mit einbeziehen –** Dies zeigt R-Quadrat- und Standardfehler-Statistiken bei jedem Zwischenschritt der							

Regression (bei Verwendung eines beliebigen Vorgangs außer Mehrfachregression).							
• **Prognose für Datensatz mit einbeziehen –** Das erstellt vorausgesagte Wer							

te für die abhängige Variable in einem zweiten Datensatz für die unabhängigen Variablen-Werte. Dieser Prognosedatensatz muss die gleichen Variablen-Namen							

wie der Originaldatensatz enthalten, der durch die Regression analysiert wird. Typischerweise sind im Prognosedatensatz Gruppen von Werten für unabhängige Vari							

ablen enthalten, für die Sie den Wert für die abhängige Variable voraussagen möchten. Um diese Prognosen vorzunehmen, wird die für den ersten Datensatz bere							

chnete Regressionsgleichung verwendet. Die vorausgesagten Werte für die abhängige Variable werden dann direkt in den Prognosedatensatz eingegeben, indem diese							

Werte einfach in die Spalte (oder Zeile) für die abhängige Variable eingefüllt werden. Die Prognose-Intervalle und Vertrauensbereiche (für die Mittelwerte) werden							

auch berücksichtigt.

Regressions-bericht

StatTools-Bericht

Ausgefü

Aktu

Mehrfachregression für Geme

Übersicht

ANOVA-Tabelle

Erklärt

Unerklärt

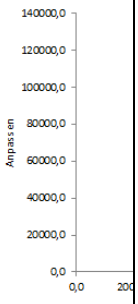
Regressionstabelle

Konstante

Maschinenstunden

Produktionsläufe

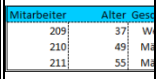
Punktdi



Datensatz mit Prognosen

Wenn die Option Prognose für Datensatz mit einbeziehe

n aktiviert ist, werden Prognosen rechts des ausgewählten Datensatzes berücksichtigt. Der Bericht enthält eine Spalte für die Prognosen, zwei Spalten für das Prognosen-Intervall und zwei Spalten für den Vertrauensbereich.



Mitarbeiter	Alter	Ges
209	37	W
210	49	Ma
211	55	Ma

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Wenn in einer Zeile irgendwel

che Wer te für eine aus gew ählt e Vari able fehl en, wir d dies e Zeil e einf ach igno riert . **Mit Daten verknüpfen** – Es ist keine Verknüpfun g mit den Originaldat en vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden. Befehl *Logistische Regression*							

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellbereichen?	Ungültige Daten	Anzahl der Var.
Befehl Diskriminanzanalyse	Führt eine Diskriminanzanalyse an einem Datensatz aus. Dabei sollte eine Kategorievariable vorhanden sein, aus der hervorgeht, in welcher von zwei oder mehr Gruppen sich die einzelnen Beobachtungen befinden; auch sollten eine oder mehrere zusätzliche Variablen verfügbar sein, um die Gruppenmitgliedschaft voraussagen zu können.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Statisch	Gestapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Nur in abhängigen Variablen zulässig	1 abhängige, 1 – 250 unabhängige
Qualitätskontrolle							
Befehl Pareto-Diagramm	Erstellt ein Pareto-Diagramm, aus dem Sie die relative Wichtigkeit der kategorisierten Daten erkennen können.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Statisch	Entstapelte Daten	Ja	Ignoriert	1 Kategorie oder 1 Wert und 1 Kategorie
Befehl x/r-Diagramme	Erstellt x- und r-Diagramme, um anzuzeigen, ob ein Vorgang statistisch erfasst wird.	Nicht zulässig	Statisch	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	2-25
Befehl p-Diagramm	Erstellt p-Diagramme, um anzuzeigen, ob ein Vorgang statistisch erfasst wird.	Nicht zulässig	Statisch	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1 Variable 1 Größen-variable
Befehl c-Diagramm	Erstellt c-Diagramme, um anzuzeigen, ob ein Vorgang statistisch erfasst wird.	Nicht zulässig	Statisch	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1
Befehl u-Diagramm	Erstellt u-Diagramme, um anzuzeigen, ob ein Vorgang statistisch erfasst wird.	Nicht zulässig	Statisch	Entstapelte Daten Bis zu 32.000 Fälle möglich	Nein	Nicht zulässig	1
Nichtparametrische Tests							
Befehl Vorzeichentest	Führt Hypothesetests für den Medianwert einer Variable oder für den Medianwert der Differenzen für ein Variablenpaar aus.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellbereichen?	Ungültige Daten	Anzahl der Var.
Befehl	Führt genau wie bei „Vorzeichentest“ die Hypothesetests aus, aber geht dabei davon aus, dass die Wahrscheinlichkeitsverteilung symmetrisch ist.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)
Befehl Mann-Whitney-Test	Führt einen Hypothesetest an zwei Werteproben aus.	Überall zulässig (d.h., sowohl zu Beginn als auch in der Mitte oder am Ende der Daten)	Echtzeit	Gestapelte und entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1 – 250 (unter Verwendung einer Werteprobe) Genau 2 (bei Verwendung von zwei oder gepaarten Werteproben)
Daten-Dienstprogramme							
Befehl Stapeln	Nimmt einen Datensatz, der separate Variablen für jede Gruppe in separaten Spalten enthält, und ermöglicht das Stapeln dieser Variablen in zwei Spalten: „Kategorie“ und „Wert“. Je nach Analysentyp könnte es vielleicht einfacher sein, mit dem gestapelten Datensatz zu arbeiten, als mit der entstapelten Version.	Ja – überall in der Variable	Statisch	Nur entstapelte Daten Bis zu 65535 Fälle möglich	Nein	n/b	1-100
Anzahl der Variablen für Analyse im gestapelten Format Wenn durch einen StatTools-Vorgang die Anzahl der für die Analyse auszuwählenden Variablen begrenzt wird, wirkt	Führt genau das Umgekehrte von „Stapeln“ aus.	Ja – überall in der Variable	n/b	Gestapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	n/b	1-32

sich diese Begrenzung auf die Anzahl der Kategorien in der Kategorien-Variable aus. Im gestapelten Format wird gewöhnlich nur eine einzige Kategorie und auch nur eine Variable für den Wert ausgewählt. Dadurch werden dann aber mehrere Variablen für die Analyse angegeben, d. h. eine für jede Kategorie in der Kategorien-Variable. Befehl *Entstapeln*							
Befehl Daten transformieren	Erstellt Pseudovariablen (0-1) auf Basis von vorhandenen Variablen.	Ja – überall in der Variable	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	n/b	1
Befehl Verzögerung	Erstellt eine neue Variable aus einem numerischen Variablenpaar, einer Kategorie und einer numerischen Variable oder aus zwei Kategorievariablen.	Ja – überall in der Variable	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Nicht zulässig	2 – 32 aus dem gleichen Datensatz

Vorgang	Beschreibung	Fehlende Daten?	Statische gegenüber Echtzeit-Berichten	Datenanforderungen	Daten aus mehreren Zellber.?	Ungült. Daten	Anzahl der Var.
Befehl Differenz	Erstellt eine neue Variable, und zwar durch Verwendung des Produkts, der Summe, des Durchschnitts, des Minimums, des Maximums oder des Min-Max-Bereichs einer oder mehrerer Variablen.	Ja – überall in der Variable	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	n/b	2 Werte-variablen oder 1 Werte- und eine Kategorie-variable oder 2 Kategorie-variablen
Befehl Interaktion	Erstellt eine neue verzögerte Variable auf Basis einer vorhandenen Variable. Eine verzögerte Variable ist einfach eine Version der Originalvariable, die sich (je nach Verzögerung) eine entsprechende Anzahl von Zeilen unterhalb des Originals befindet.	Ja – überall in der Variable	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1
Befehl Kombination	Wendet eine der folgenden vier nicht linearen Transformationen – Natürlicher Logarithmus, Quadrat, Quadratwurzel oder Wechselseitig – auf die ausgewählten Variablen an, um eine neue Variable zu erstellen.	Ja – überall in der Variable	Echtzeit oder statisch	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-100
Befehl Pseudo	Erstellt eine beliebige Anzahl von Differenzvariablen auf Basis einer Originalvariable.	Ja – überall in der Variable	Echtzeit	Nur entstapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	n/b	1
Befehl Zufallsprobe	Ermöglicht Ihnen, jede beliebige Anzahl an Zufallswerten aus einem gegebenen Datensatz zu erstellen, wobei es keine Rolle spielt, ob die Probenerhebung mit oder ohne Ersetzung ist.	Ja – überall in der Variable	Statisch	Gestapelte Daten Bis zu 16 Millionen Fälle möglich	Ja	Ignoriert	1-32

StatTools-Menü – Datensätze

Befehl *Datensatzmanager*

Definiert StatTools-Datensätze und Variablen oder bearbeitet bzw. löscht einen bestehenden Datensatz nebst Variablen

Der Befehl **Datensatzmanager** ermöglicht Ihnen, die Datensätze und Variablen zu definieren. Sobald die Datensätze und Variablen definiert sind, können Sie in StatTools analysiert werden. Über das Dialogfeld **Datensatzmanager** können Sie Datensätze hinzufügen oder entfernen sowie auch einen Datensatz benennen oder das Layout der Variablen in einem Datensatz angeben und diese Variablen wie gewünscht bezeichnen.

Was sind Datensätze und Variablen?

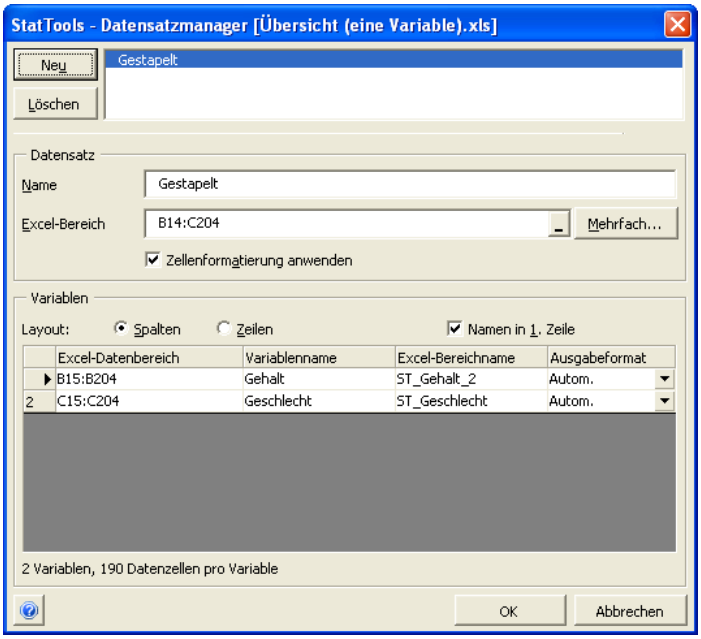
StatTools funktioniert so ähnlich wie die meisten eigenständigen statistischen Softwarepakete, indem es den Variablen gemäß strukturiert ist. Bei den meisten Analysen ist es erforderlich, dass Sie mit einem Datensatz oder einem Satz von statistischen Variablen arbeiten, die sich oft in zusammenhängenden Spalten mit verschiedenen Namen befinden. Die Namen der Variablen sind gewöhnlich in der ersten Zeile des Datensatzes zu sehen. Diese vordefinierten Variablen können dann beim Ausführen von Statistikanalysen verwendet werden, und zwar ohne dass die zu analysierenden Daten immer wieder neu ausgewählt werden müssen.

Jede in einem Datensatz befindliche Variable hat einen Namen und ist in Excel mit einem bestimmten Zellenbereich verknüpft. Durch das ausgewählte **Layout** wird angegeben, wie Variablen in einem Datensatz positioniert werden sollen. Gewöhnlich sind **Spalten** mit jeweils einer Variable pro Spalte vorhanden, aber für die Variablen können auch Zeilen eingerichtet werden. Ein Datensatz kann aus mehreren Zellblöcken bestehen, wodurch Sie dann Daten in verschiedene Blätter derselben Arbeitsmappe eingeben können.

Wenn Sie einen Datensatz definieren, versucht StatTools, in Excel die Variablen zu identifizieren, die sich in den Zellen um die aktuelle Auswahl befinden. Dadurch kann schnell und mühelos ein Datensatz eingerichtet werden, und zwar mit den Variablen in Spalten und den entsprechenden Namen der Variablen jeweils in der obersten Zeile.

In Ihrem Datensatz brauchen die Spalten nicht unbedingt alle gleich lang zu sein. Sie könnten beispielsweise zwei Variablen – *Gewicht_Männer* und *Gewicht_Frauen* – mit unterschiedlicher Anzahl von Beobachtungen haben. Bei vielen Analysen wird StatTools jedoch die leeren Zellen in den kürzeren Spalten einfach als fehlende Daten behandeln.

Dialogfeld
Datensatzmanager



Im Dialogfeld **Datensatzmanager** sind unter **Datensatz** folgende Optionen zu finden:

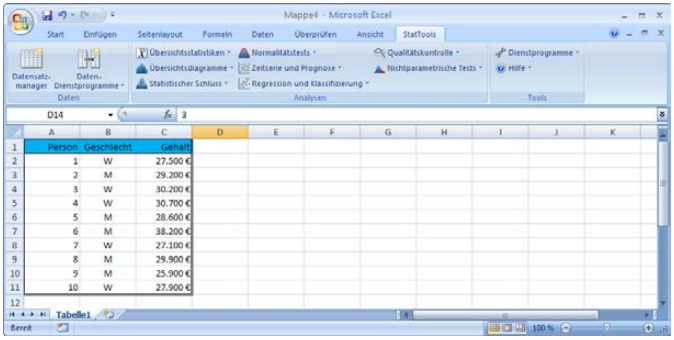
- **Neu, Löschen** – fügt einen neuen Datensatz hinzu oder löscht einen bereits vorhandenen
- **Name** – legt den Namen für den Datensatz fest
- **Excel-Bereich** – bestimmt den mit einem Datensatz verknüpften Excel-Bereich Wenn einem Datensatz mehrere Zellbereiche zugewiesen wurden, steht vor diesem Eintrag die Bezeichnung **Mehrfach**.
- **Zellformatierung anwenden** – fügt Raster und Farben hinzu, um die Datensätze zu identifizieren
- Mehrfach – durch Klicken auf **Mehrfach** im Dialogfeld **Datensatzmanager** wird das Dialogfeld **Mehrfachbereichs-Selektor** angezeigt. Dieses Dialogfeld ermöglicht Ihnen, die einzelnen Zellbereiche einzugeben, aus denen sich der aus mehreren Zellbereichen bestehende Datensatz zusammensetzt.

- **Sekundäre Bereiche haben die Variablennamen in der ersten Zeile –** Bei Mehrfachbereichs-Datensätzen können sich die Namen von Variablen oben in jeder Spalte (oder, je nach ausgewähltem Variablen-Layout, auch in jeder Zeile) der einzelnen im Dialogfeld aufgelisteten Bereiche befinden. Es ist aber auch möglich, dass die Namen der Variablen nur die Spalten oder Zeilen im ersten ausgewählten Bereich bezeichnen. Der erste ausgewählte Bereich ist der Bereich, der in Zeile 1 des Dialogfelds **Mehrfachbereichs-Selektor** eingegeben wurde.

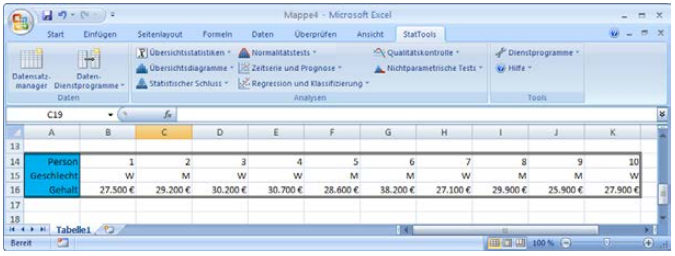
Im Dialogfeld **Datensatzmanager** sind unter **Variablen** folgende Optionen zu finden:

- **Layout –** kennzeichnet, wie die Variablen im Excel-Bereich (in dem sich der Datensatz befindet) strukturiert sind. Für **Layout** sind folgende Optionen verfügbar:

  - **Spalten**. Dies ist das typische Layout, in dem jede Spalte im Excel-Bereich des Datensatzes die Daten für eine Variable enthält. Die Namen der Variablen werden oft oben in die einzelnen Spalten eingegeben.



  - **Zeilen**. In diesem Layout enthält jede Zeile des Datensatzes die Daten für eine Variable. Dieses Layout wird in Excel oft für Zeitseriendaten verwendet.

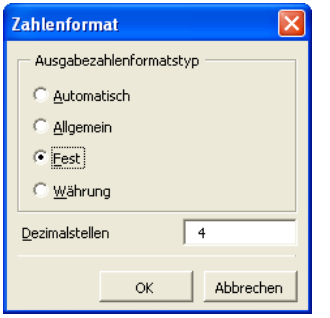


- **Namen in 1. Spalte (oder Zeile)** – Wählen Sie diese Option, wenn in einem Datensatz oben in den Spalten (oder bei **Zeilen**-Layout in den ganz links befindlichen Zellen) die Namen der Variablen angegeben sind.

Optionen für Variablen

Im Dialogfeld **Datensatzmanager** sind in jeder Rasterzeile die Variablen eines Datensatzes aufgelistet, einschließlich des Namens der einzelnen Variablen, des Excel-Bereichs, der die Datenpunkte für die Variable enthält, und des Excel-Bereichsnamens, der dazu verwendet wird, in Excel-Formeln die Daten für die Variable zu identifizieren.

- **Excel-Bereichsname** – Der angezeigte Bereichsname wird in den in StatTools-Berichten und –Diagrammen erstellten Excel-Formeln verwendet. Dadurch können die Berichte in Echtzeit, d.h., automatisch aktualisiert werden, sobald sich die Daten einer Variable ändern. Durch für Sie verständliche Bereichsnamen können die Formeln lesbarer gemacht werden.
- **Ausgabeformat** – Hierdurch wird das Format der Werte angegeben, die in StatTools-Analyseberichten für eine Variable angezeigt werden. Durch den Eintrag **Automatisch** wird angegeben, dass StatTools das **beste** Format auswählen soll, und zwar auf Basis der numerischen Formatierung, die in Excel auf die Zellen angewendet wird, die die Werte der Variable enthalten. Durch Klicken auf den Pfeil, der sich neben dem Eintrag **Ausgabeformat** befindet, können Sie aber auch ein bestimmtes von Ihnen gewünschtes Format auswählen:



Allgemein entspricht dem allgemeinen numerischen Format in Excel. Durch **Fest** wird die Genauigkeit angegeben, und zwar unter Verwendung der eingegebenen Anzahl an **Dezimalstellen**. **Währung** entspricht dem numerischen Format **Währung** in Excel und zeigt die Genauigkeit, indem die eingegebene Anzahl an **Dezimalstellen** verwendet wird.

Hinweis: Das gewünschte Ausgabeformat kann auch direkt in das Dialogfeld **Datensatzmanager** eingegeben werden, und zwar unter Verwendung der Notation *selectedFormat(#decimalDigits)*, such as *Währung(4)*.

Fähigkeiten von Datensatz und Variablen

In einer einzigen Sitzung ermöglicht StatTools Folgendes:

- bis zu 256 Datensätzen in einer Arbeitsmappe
- bis zu 256 Variablen pro Datensatz; alle Daten für einen Datensatz müssen sich in derselben Arbeitsmappe befinden
- bis zu 16.777.216 Datenpunkte pro Variable

Die wirklichen Datenkapazitäten sind je nach Systemkonfiguration und verwendeter Excel-Version evtl. geringer als die vorstehend angegebenen. Bestimmte StatTools-Analysen können verschiedene Beschränkungen oder Begrenzungen haben. Die Speicherbegrenzungen von Excel können sich ebenfalls auf die Datenkapazitäten auswirken.

Hinweis: Im Dialogfeld* Datensatzmanager *sind alle in der aktiven Arbeitsmappe befindlichen Datensätze und Variablen aufgelistet (es handelt sich dabei um die oben im Dialogfeld* Datensatzmanager *angegebene Arbeitsmappe). Um in anderen Arbeitsmappen befindliche Datensätze auflisten zu können, muss die gewünschte Arbeitsmappe in Excel aktiviert und dann im Dialogfeld* Datensatzmanager *angezeigt werden.

Menü *Dienstprogramme*

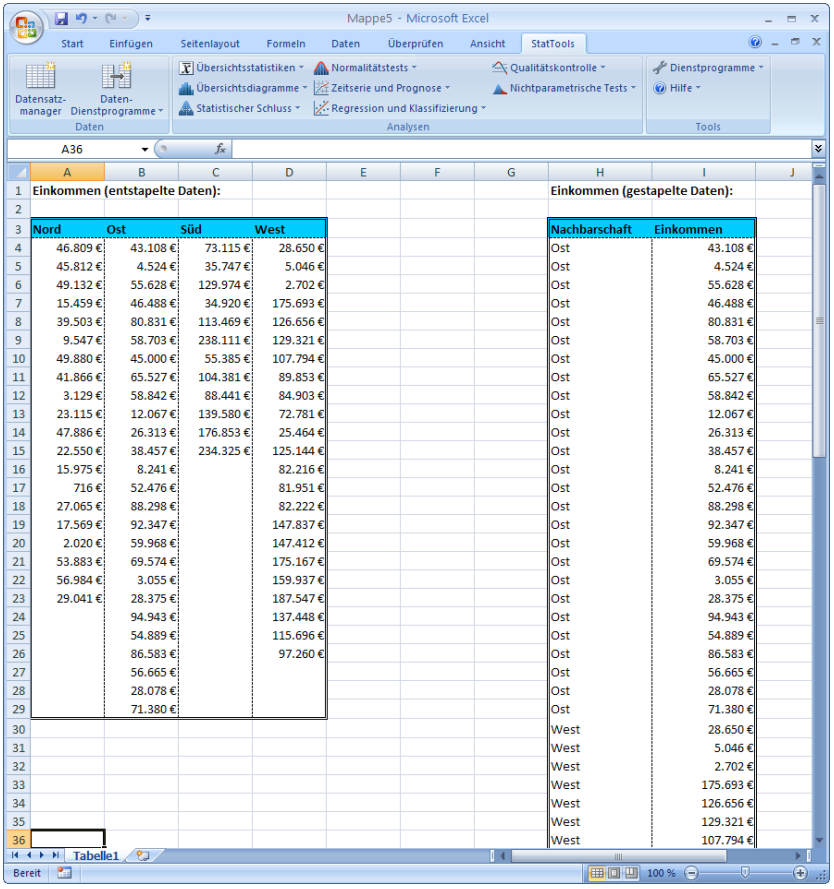
Befehl *Stapeln*

Konvertiert einen Satz von entstapelten Variablen in das gestapelte Format

Der Befehl **Stapeln** ermöglicht Ihnen, **ungestapelte** Daten aus einem Datensatz mit mindestens zwei *Wert*variablen in ein **gestapeltes** Format zu konvertieren, in dem der Datensatz dann eine *Kategorie*- und eine *Wert*variable enthält. Wenn wir beispielsweise die durchschnittlichen Haushaltseinkommen in mehreren unterschiedlichen Nachtbarschaften vergleichen, würde bei der entstapelten Form eine separate *Einkommen*-Variable (oder Spalte) für jede Nachtbarschaft vorhanden sein. Diese Spalten brauchen nicht von gleicher Länge sein, d.h., für jede Nachbarschaft könnte eine unterschiedliche Probengröße verwendet werden. Bei der gestapelten Form würde dagegen die Wertevariable *Einkommen* und die Kategorievariable *Nachbarschaft* verfügbar sein, um zu erkennen zu geben, in welcher Nachbarschaft sich die einzelnen Haushalte befinden.

Im Wesentlichen ermöglicht Ihnen dieses Verfahren, Ihre Variablen in zwei Spalten zu stapeln, und zwar in der Wertespalte *Einkommen* und der Kategoriespalte *Nachbarschaft*. Je nach Analysentyp könnte es vielleicht einfacher sein, mit dem gestapelten Datensatz zu arbeiten, als mit der entstapelten Version.

Gestapelte und entstapelte Variablen



Mappe5 - Microsoft Excel

Einkommen (entstapelte Daten):

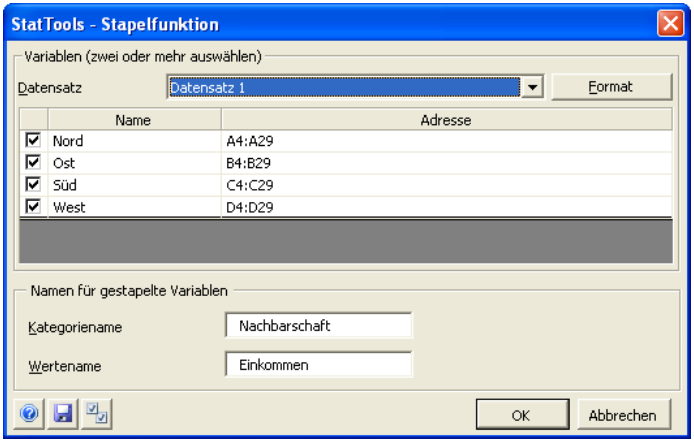
Nord	Ost	Süd	West
46.809 €	43.108 €	73.115 €	28.650 €
45.812 €	4.524 €	35.747 €	5.046 €
49.132 €	55.628 €	129.974 €	2.702 €
15.459 €	46.488 €	34.920 €	175.693 €
39.503 €	80.831 €	113.469 €	126.656 €
9.547 €	58.703 €	238.111 €	129.321 €
49.880 €	45.000 €	55.385 €	107.794 €
41.866 €	65.527 €	104.381 €	89.853 €
3.129 €	58.842 €	88.441 €	84.903 €
23.115 €	12.067 €	139.580 €	72.781 €
47.886 €	26.313 €	176.853 €	25.464 €
22.550 €	38.457 €	234.325 €	125.144 €
15.975 €	8.241 €		82.216 €
716 €	52.476 €		81.951 €
27.065 €	88.298 €		82.222 €
17.569 €	92.347 €		147.837 €
2.020 €	59.968 €		147.412 €
53.883 €	69.574 €		175.167 €
56.984 €	3.055 €		159.937 €
29.041 €	28.375 €		187.547 €
	94.943 €		137.448 €
	54.889 €		115.696 €
	86.583 €		97.260 €
	56.665 €		
	28.078 €		
	71.380 €		

Einkommen (gestapelte Daten):

Nachbarschaft	Einkommen
Ost	43.108 €
Ost	4.524 €
Ost	55.628 €
Ost	46.488 €
Ost	80.831 €
Ost	58.703 €
Ost	45.000 €
Ost	65.527 €
Ost	58.842 €
Ost	12.067 €
Ost	26.313 €
Ost	38.457 €
Ost	8.241 €
Ost	52.476 €
Ost	88.298 €
Ost	92.347 €
Ost	59.968 €
Ost	69.574 €
Ost	3.055 €
Ost	28.375 €
Ost	94.943 €
Ost	54.889 €
Ost	86.583 €
Ost	56.665 €
Ost	28.078 €
Ost	71.380 €
West	28.650 €
West	5.046 €
West	2.702 €
West	175.693 €
West	126.656 €
West	129.321 €
West	107.794 €

Dialogfeld Stapelfunktion

Zum Stapeln von Variablen wird das Dialogfeld **Stapelfunktion** verwendet:



Um stapeln zu können, müssen mindestens zwei Variablen ausgewählt werden. Vom ausgewählten Datensatz wird anfänglich immer angenommen, dass es sich um entstapelte Daten handelt. Variablen können aus verschiedenen Datensätzen stammen.

Das Dialogfeld **Stapelfunktion** enthält die

- **Namen für gestapelte Variablen** – Dabei handelt es sich um die Namen der *Kategorie*- und *Wert*variable, aus denen sich der aus zwei Variablen bestehende gestapelte Datensatz zusammensetzt. Diese Namen erscheinen ganz oben in den Spalten, die für die *Kategorie*- und für die *Wert*variablen vorgesehen sind.

Sobald Sie auf OK klicken, werden die Variablen gestapelt und wird dann für die gestapelten Daten ein neuer Datensatz erstellt.

Was sind Kategorie- und Wertvariablen?

Diese Variablen sind für einen gestapelten Datensatz erforderlich. Die *Kategorie*variable (die mitunter auch *Code*variable genannt wird) ist einfach eine beschreibende Identifizierung des damit verbundenen Satzes von *Wert*variablen. Bei der *Kategorie*variable handelt es sich oft um eine Textbezeichnung. *Wert*variablen (die mitunter auch als *Messwert*variablen bezeichnet werden) sind dagegen standardmäßige numerische Variablen, die durch statistische Vorgänge analysiert werden können.

Anzahl der Variablen für Analyse im gestapelten Format

Wenn durch einen StatTools-Vorgang die Anzahl der für die Analyse auszuwählenden Variablen begrenzt wird, wirkt sich diese Begrenzung auf die Anzahl der Kategorien in der Kategorien-Variable aus. Im gestapelten Format wird gewöhnlich nur eine einzige Kategorie und auch nur eine Variable für den Wert ausgewählt. Dadurch werden dann aber mehrere Variablen für die Analyse angegeben, d. h. eine für jede Kategorie in der Kategorien-Variable.

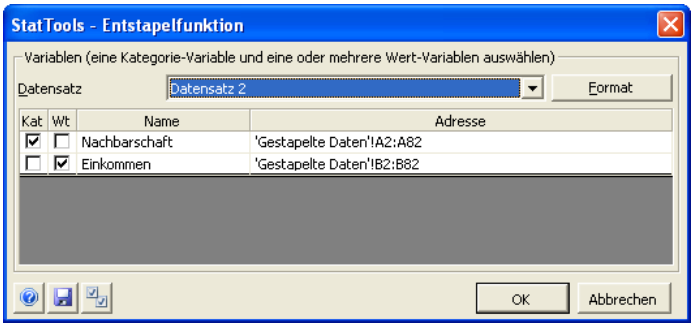
Befehl *Entstapeln*

Konvertiert einen Satz von gestapelten Variablen in das entstapelte Format

Durch den Befehl **Entstapeln** wird genau das Entgegengesetzte vom Befehl **Stapeln** ausgeführt. Wenn Sie beispielsweise mit der Kategorievariable *Geschlecht* und der Wertvariable *Gewicht* beginnen, werden die beiden Variablen durch diesen Befehl in die separaten Spalten *Männergewicht* und *Frauengewicht* entstapelt.

***Dialogfeld* Entstapelfunktion**

Zum Entstapeln von Variablen wird das Dialogfeld **Entstapelfunktion** verwendet:



Vom ausgewählten Datensatz wird hier anfänglich immer angenommen, dass es sich um gestapelte Daten handelt. Um entstapeln zu können, müssen mindestens zwei Variablen ausgewählt werden. Eine dieser Variablen wird als *Kategorie*variable angegeben (durch Häkchen im Kontrollkästchen **Kat**.) und die andere oder anderen Variablen als *Wert*variablen (durch Häkchen im Kontrollkästchen **Wert**). Variablen können aus verschiedenen Datensätzen stammen.

Sobald Sie auf OK klicken, werden die Variablen entstapelt und wird dann für die entstapelten Daten ein neuer Datensatz erstellt.

Befehl *Daten transformieren*

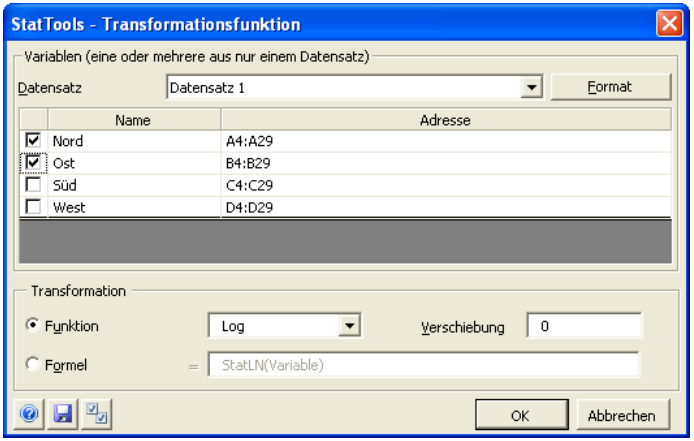
Formt eine oder mehrere Variablen in neue Variablen und Werte um, und zwar auf Basis einer eingegebenen Umformungsfunktion

Mithilfe des Befehls **Daten transformieren** können Sie jede beliebige Variable auf vier verschiedene Weisen umformen: **Natürlicher Logarithmus, Quadrat, Quadratwurzel** oder **Wechselseitig**. Auch ermöglicht Ihnen dieser Befehl, eine Formel einzugeben, durch die dann ein Wert für die umgeformte Variable berechnet wird.

Wenn für die Variable, auf der die umgeformte Variable basiert, nicht alle Werte vorhanden sind, werden die entsprechenden Werte auch in der umgeformten Variable fehlen.

***Dialogfeld* Transformations-funktion**

Zum Umformen oder Transformieren von Variablen wird das Dialogfeld **Transformationsfunktion** verwendet:



Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Es können jeweils nur Variablen aus demselben Datensatz umgeformt werden.

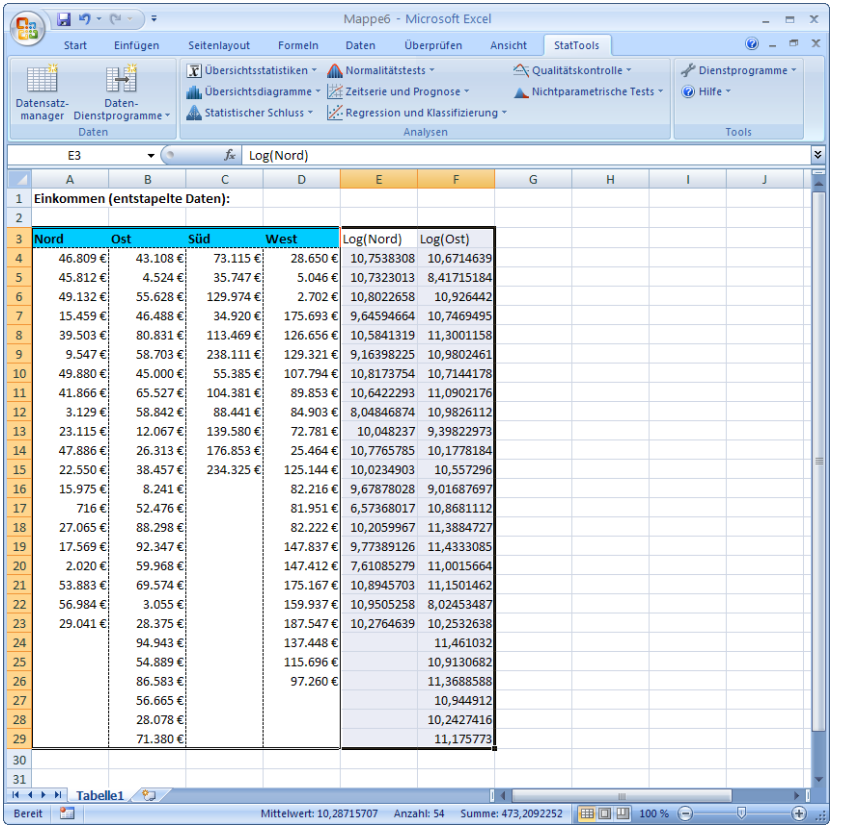
Das Dialogfeld **Transformationsfunktion** enthält die

- **Transformationsfunktion –** d.h., den mathematischen Vorgang, der beim Erstellen des neuen umgeformten Wertes an jedem Wert der ausgewählten Variablen ausgeführt wird. Bei den integrierten Transformationsfunktionen handelt es sich um die Funktionen **Natürlicher Logarithmus, Quadrat, Quadratwurzel** und **Wechselseitig.** Auch kann eine benutzerdefinierte **Formel** eingegeben werden, durch die ein neuer variablen Wert berechnet werden kann, und zwar auf Basis eines mathematischen Ausdrucks, wie z.B.:

  *(Variable*1.5)^2*

  In dieser Gleichung wird das Schlüsselwort „Variable“ als Platzhalter für den effektiven Wert der umzuformenden Variable verwendet.

Datensatz mit neuen umgeformten Variablen



	A	B	C	D	E	F
1	Einkommen (entstapelte Daten):					
2						
3	Nord	Ost	Süd	West	Log(Nord)	Log(Ost)
4	46.809 €	43.108 €	73.115 €	28.650 €	10,7538308	10,6714639
5	45.812 €	4.524 €	35.747 €	5.046 €	10,7323013	8,41715184
6	49.132 €	55.628 €	129.974 €	2.702 €	10,8022658	10,926442
7	15.459 €	46.488 €	34.920 €	175.693 €	9,64594664	10,7469495
8	39.503 €	80.831 €	113.469 €	126.656 €	10,5841319	11,3001158
9	9.547 €	58.703 €	238.111 €	129.321 €	9,16398225	10,9802461
10	49.880 €	45.000 €	55.385 €	107.794 €	10,8173754	10,7144178
11	41.866 €	65.527 €	104.381 €	89.853 €	10,6422293	11,0902176
12	3.129 €	58.842 €	88.441 €	84.903 €	8,04846874	10,9826112
13	23.115 €	12.067 €	139.580 €	72.781 €	10,048237	9,39822973
14	47.886 €	26.313 €	176.853 €	25.464 €	10,7765785	10,1778184
15	22.550 €	38.457 €	234.325 €	125.144 €	10,0234903	10,557296
16	15.975 €	8.241 €		82.216 €	9,67878028	9,01687697
17	716 €	52.476 €		81.951 €	6,57368017	10,8681112
18	27.065 €	88.298 €		82.222 €	10,2059967	11,3884727
19	17.569 €	92.347 €		147.837 €	9,77389126	11,4333085
20	2.020 €	59.968 €		147.412 €	7,61085279	11,0015664
21	53.883 €	69.574 €		175.167 €	10,8945703	11,1501462
22	56.984 €	3.055 €		159.937 €	10,9505258	8,02453487
23	29.041 €	28.375 €		187.547 €	10,2764639	10,2532638
24		94.943 €		137.448 €		11,461032
25		54.889 €		115.696 €		10,9130682
26		86.583 €		97.260 €		11,3688588
27		56.665 €				10,944912
28		28.078 €				10,2427416
29		71.380 €				11,175773
30						
31						

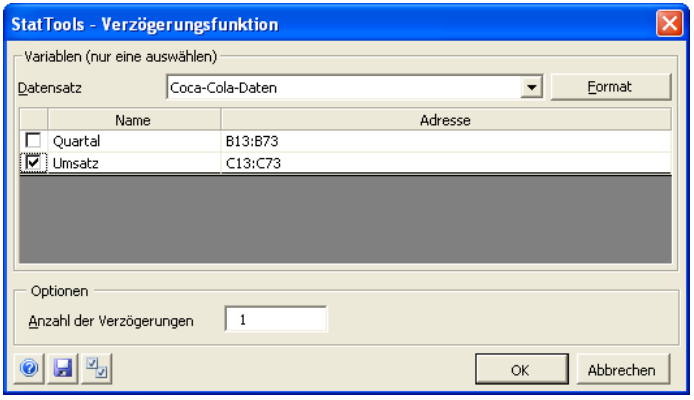
Befehl *Verzögerung*

Erstellt eine neue verzögerte Variable auf Basis einer bereits vorhandenen Variable

Der Befehl **Verzögerung** ermöglicht Ihnen, auf Basis einer bereits vorhandenen Variable eine neue verzögerte Variable zu erstellen. Eine verzögerte Variable ist einfach eine Version der Originalvariable, die sich (je nach Verzögerung) eine entsprechende Anzahl von Zeilen unterhalb des Originals befindet Die um 3 (Monate) verzögerte Version der Verkäufe für November 1998 bezieht sich beispielsweise auf die drei Monate früheren Verkäufe im August 1998.

***Dialogfeld* Verzögerungs-funktion**

Zum Verzögern von Variablen wird das Dialogfeld **Verzögerungsfunktion** verwendet:

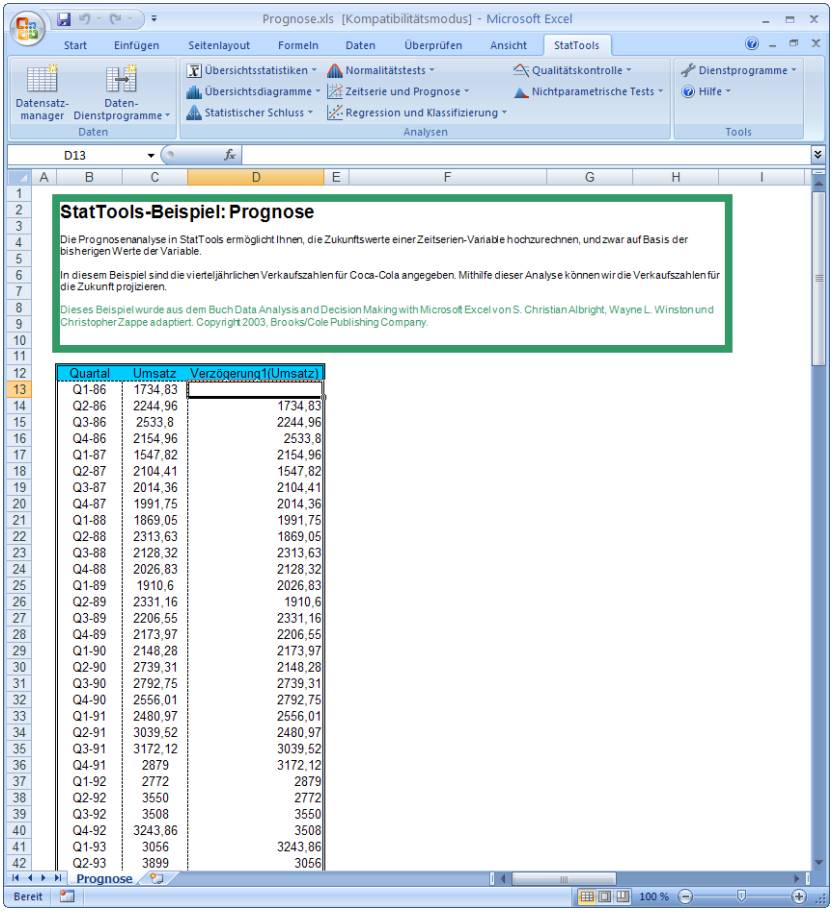


Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Es kann nur jeweils eine Variable verzögert werden.

In diesem Dialogfeld finden Sie unter **Optionen** Folgendes:

- **Anzahl der Verzögerungen** – oder wie oft Werte beim Erstellen von neuen Variablen verzögert werden sollen. Bei jeder Verzögerung (bis hin zur eingegebenen Verzögerungsanzahl) wird eine neue Variable erstellt.

Verzögerte Variable rechts vom Quelldatensatz



Quartal	Umsatz	Verzögerung1(Umsatz)
Q1-86	1734,83	
Q2-86	2244,96	1734,83
Q3-86	2533,8	2244,96
Q4-86	2154,96	2533,8
Q1-87	1547,82	2154,96
Q2-87	2104,41	1547,82
Q3-87	2014,36	2104,41
Q4-87	1991,75	2014,36
Q1-88	1869,05	1991,75
Q2-88	2313,63	1869,05
Q3-88	2128,32	2313,63
Q4-88	2026,83	2128,32
Q1-89	1910,6	2026,83
Q2-89	2331,16	1910,6
Q3-89	2206,55	2331,16
Q4-89	2173,97	2206,55
Q1-90	2148,28	2173,97
Q2-90	2739,31	2148,28
Q3-90	2792,75	2739,31
Q4-90	2556,01	2792,75
Q1-91	2480,97	2556,01
Q2-91	3039,52	2480,97
Q3-91	3172,12	3039,52
Q4-91	2879	3172,12
Q1-92	2772	2879
Q2-92	3550	2772
Q3-92	3508	3550
Q4-92	3243,86	3508
Q1-93	3056	3243,86
Q2-93	3899	3056

Befehl *Differenz*

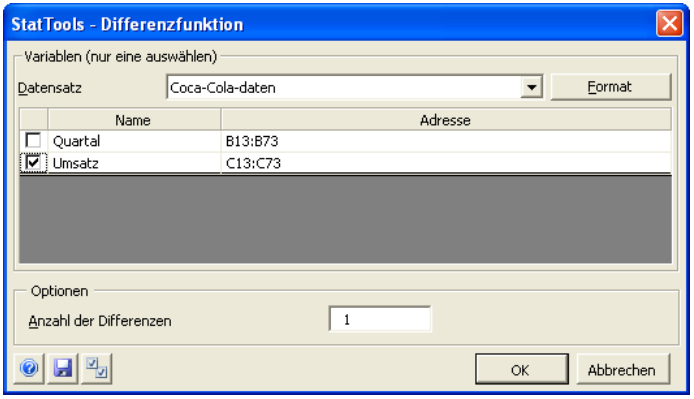
Erstellt eine beliebige Anzahl von Differenzvariablen auf Basis einer Originalvariable

Mithilfe dieses Befehls können Sie jede beliebige Anzahl von Differenzvariablen aus einer Originalvariable erstellen. Der Befehl **Differenz** wird hauptsächlich für Zeitserienvariablen verwendet. Die zu differenzierende Variable wird zusammen mit der Anzahl der Differenzen (in der Regel 1 oder 2) ausgewählt. Durch die Funktion werden dann entsprechend viele neue Differenzvariablen erstellt. Jede dieser Differenzvariablen differenziert sich in bestimmter Weise von der ausgewählten Variable. Bei monatlichen Daten besteht der Differenzwert für März 1997 beispielsweise aus dem Originalwert für März 1997 minus dem Originalwert für Februar 1997. In ähnlicher Weise enthält die zweite Differenzvariable (falls benötigt) den Unterschied zwischen den ersten Differenzen.

Das Differenzieren wird oft in der Zeitserienanalyse verwendet, wenn die Originalvariable mit fortlaufend „stationär" ist. Eine Zeitserie mit steigender Tendenz ist beispielsweise nicht stationär. Durch das Differenzieren wird oft die Stationarität erreicht. Mitunter ist eine zweite Differenzierung nützlich, was aber nur seltener der Fall ist. Dritte oder noch darüber hinausgehende Differenzierungen sind fast nie erforderlich.

***Dialogfeld* Differenzfunktion**

Zum Erstellen von Differenzvariablen wird das Dialogfeld **Differenzfunktion** verwendet:

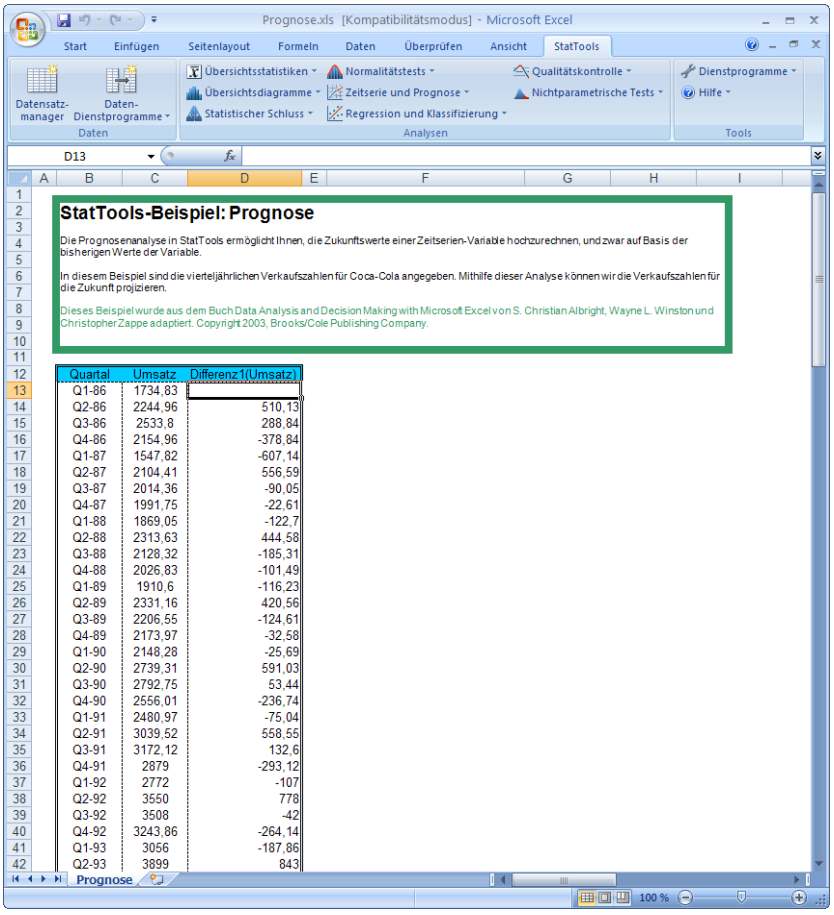


Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Zum Erstellen von Differenzvariablen kann nur jeweils eine Variable verwendet werden.

In diesem Dialogfeld finden Sie unter **Optionen** Folgendes:

- **Anzahl der Differenzen** – oder Anzahl der zu erstellenden Differenzen

Datensatz mit Differenzvariable



Quartal	Umsatz	Differenz1(Umsatz)
Q1-86	1734,83	
Q2-86	2244,96	510,13
Q3-86	2533,8	288,84
Q4-86	2154,96	-378,84
Q1-87	1547,82	-607,14
Q2-87	2104,41	556,59
Q3-87	2014,36	-90,05
Q4-87	1991,75	-22,61
Q1-88	1869,05	-122,7
Q2-88	2313,63	444,58
Q3-88	2128,32	-185,31
Q4-88	2026,83	-101,49
Q1-89	1910,6	-116,23
Q2-89	2331,16	420,56
Q3-89	2206,55	-124,61
Q4-89	2173,97	-32,58
Q1-90	2148,28	-25,69
Q2-90	2739,31	591,03
Q3-90	2792,75	53,44
Q4-90	2556,01	-236,74
Q1-91	2480,97	-75,04
Q2-91	3039,52	558,55
Q3-91	3172,12	132,6
Q4-91	2879	-293,12
Q1-92	2772	-107
Q2-92	3550	778
Q3-92	3508	-42
Q4-92	3243,86	-264,14
Q1-93	3056	-187,86
Q2-93	3899	843

Befehl *Interaktion*

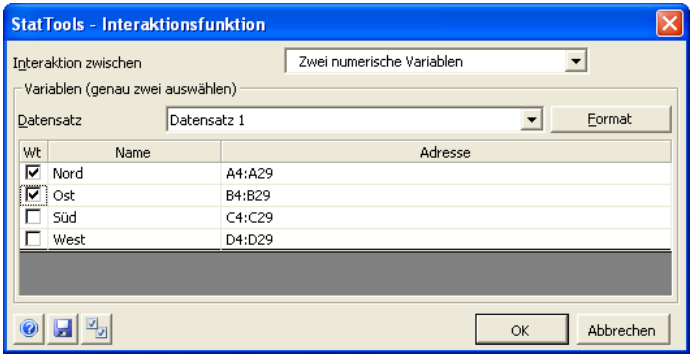
Erstellt eine Interaktionsvariable aus einer oder mehreren Originalvariablen

Dieser Befehl ermöglicht Ihnen, eine Interaktionsvariable aus einer oder mehreren Originalvariablen zu erstellen. Zum Erstellen von Interaktionsvariablen können zwei numerische Variablen, eine numerische und eine Kategorievariable oder auch zwei Kategorievariablen verwenden werden.

Falls beide Variablen numerisch (und nicht kategorisch) sind, wird daraus das Produkt der beiden Variablen erstellt. Wenn eine Variable numerisch und die andere kategorisch ist, werden daraus die Produkte der numerischen Variable erstellt und jede Pseudovariable entspricht dann den Kategorien der kategorischen Variable. Wenn dagegen beide Variablen kategorisch sind, werden die Produkte aller Pseudopaare aus den beiden kategorischen Variablen erstellt.

***Dialogfeld* Interaktionsfunktion**

Zum Erstellen von Interaktionsvariablen wird das Dialogfeld **Interaktionsfunktion** verwendet:



Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Es können jeweils eine oder mehrere Variablen zum Erstellen einer Interaktionsvariable verwendet werden.

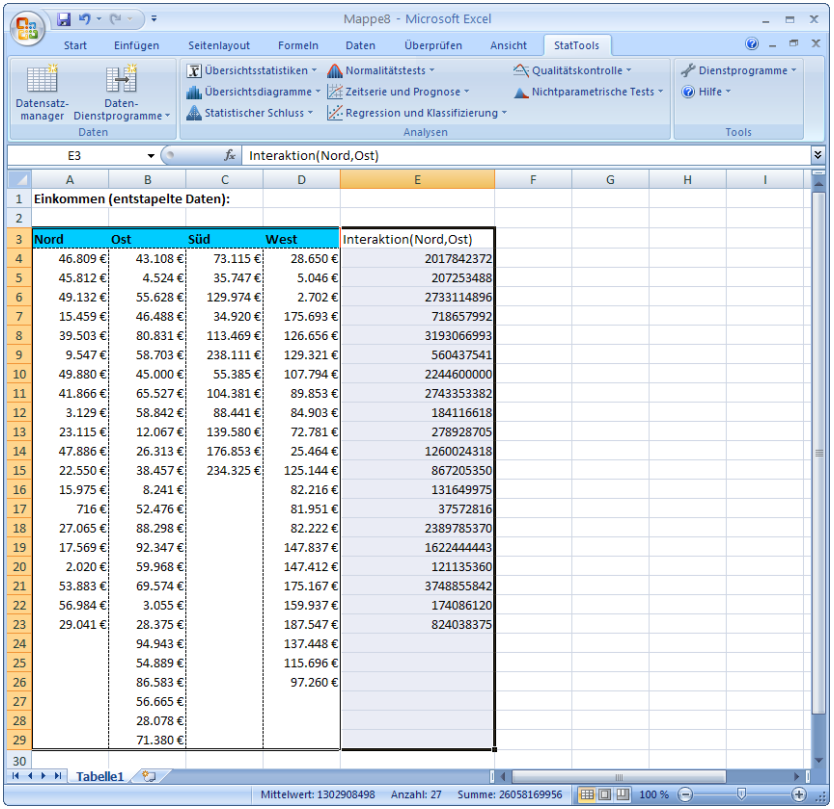
In diesem Dialogfeld finden Sie Folgendes:

- **Interaktion zwischen** – wählt den Typ der beiden auszuwählenden Variablen aus. Es kann sich dabei um **zwei numerische Variablen, eine numerische und eine Kategorievariable** oder um **zwei Kategorievariablen** handeln.

Wie wird eine Interaktionsvariable erstellt?

Eine Interaktionsvariable wird aus zwei Variablen erstellt, die Sie im Dialogfeld auswählen. Für diese beiden Variablen sind drei grundlegende Optionen vorhanden. Erstens kann es sich bei beiden um numerische „Messwert“-Variablen handeln. In diesem Fall besteht die Interaktionsvariable dann aus dem Produkt dieser beiden Variablen. Zweitens können eine numerische Messwertvariable und eine Kategorievariable verwendet werden. Dadurch werden in StatTools interne Pseudovariablen für jede Kategorie der Kategorievariable erstellt und werden die einzelnen Pseudos dann mit der numerischen Variable multipliziert. Drittens können auch beide Variablen Kategorievariablen sein. Dadurch werden in StatTools interne Pseudovariablen für jede Kategorie der beiden Kategorievariablen erstellt und werden die Pseudos der ersten Variable dann mit den Pseudos der zweiten multipliziert. Wenn die beiden Kategorievariablen beispielsweise 2 bzw. 5 Kategorien aufweisen, erstellt StatTools 2 x 5 = 10 Interaktionsvariablen.

Datensatz mit Interaktionsvariable



	A	B	C	D	E
1	Einkommen (entstapelte Daten):				
2					
3	Nord	Ost	Süd	West	Interaktion(Nord,Ost)
4	46.809 €	43.108 €	73.115 €	28.650 €	2017842372
5	45.812 €	4.524 €	35.747 €	5.046 €	207253488
6	49.132 €	55.628 €	129.974 €	2.702 €	2733114896
7	15.459 €	46.488 €	34.920 €	175.693 €	718657992
8	39.503 €	80.831 €	113.469 €	126.656 €	3193066993
9	9.547 €	58.703 €	238.111 €	129.321 €	560437541
10	49.880 €	45.000 €	55.385 €	107.794 €	2244600000
11	41.866 €	65.527 €	104.381 €	89.853 €	2743353382
12	3.129 €	58.842 €	88.441 €	84.903 €	184116618
13	23.115 €	12.067 €	139.580 €	72.781 €	278928705
14	47.886 €	26.313 €	176.853 €	25.464 €	1260024318
15	22.550 €	38.457 €	234.325 €	125.144 €	867205350
16	15.975 €	8.241 €		82.216 €	131649975
17	716 €	52.476 €		81.951 €	37572816
18	27.065 €	88.298 €		82.222 €	2389785370
19	17.569 €	92.347 €		147.837 €	1622444443
20	2.020 €	59.968 €		147.412 €	121135360
21	53.883 €	69.574 €		175.167 €	3748855842
22	56.984 €	3.055 €		159.937 €	174086120
23	29.041 €	28.375 €		187.547 €	824038375
24		94.943 €		137.448 €	
25		54.889 €		115.696 €	
26		86.583 €		97.260 €	
27		56.665 €			
28		28.078 €			
29		71.380 €			
30					

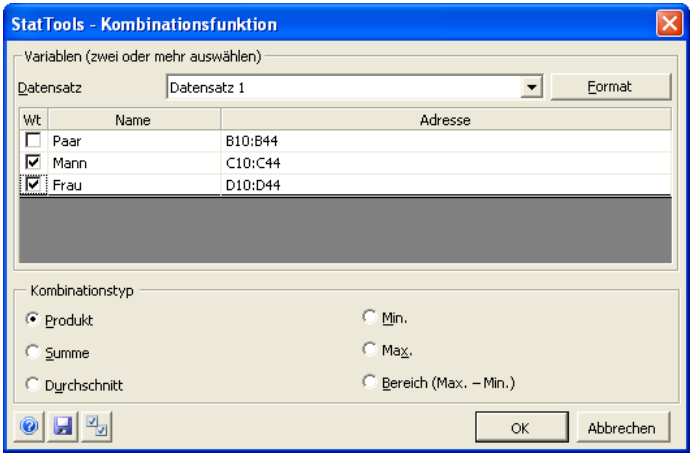
Befehl *Kombination*

Erstellt eine Kombinationsvariable aus einer oder mehreren Originalvariablen

Dieser Befehl ermöglicht Ihnen, eine Kombinationsvariable aus einer oder mehreren Originalvariablen zu erstellen. Erstellt eine Kombinationsvariable, und zwar durch Verwendung des Produkts, der Summe, des Durchschnitts, des Minimums, des Maximums oder des Min-Max-Bereichs einer oder mehrerer Variablen.

***Dialogfeld* Kombinations-funktion**

Zum Erstellen von Kombinationsvariablen wird das Dialogfeld **Kombinationsfunktion** verwendet:

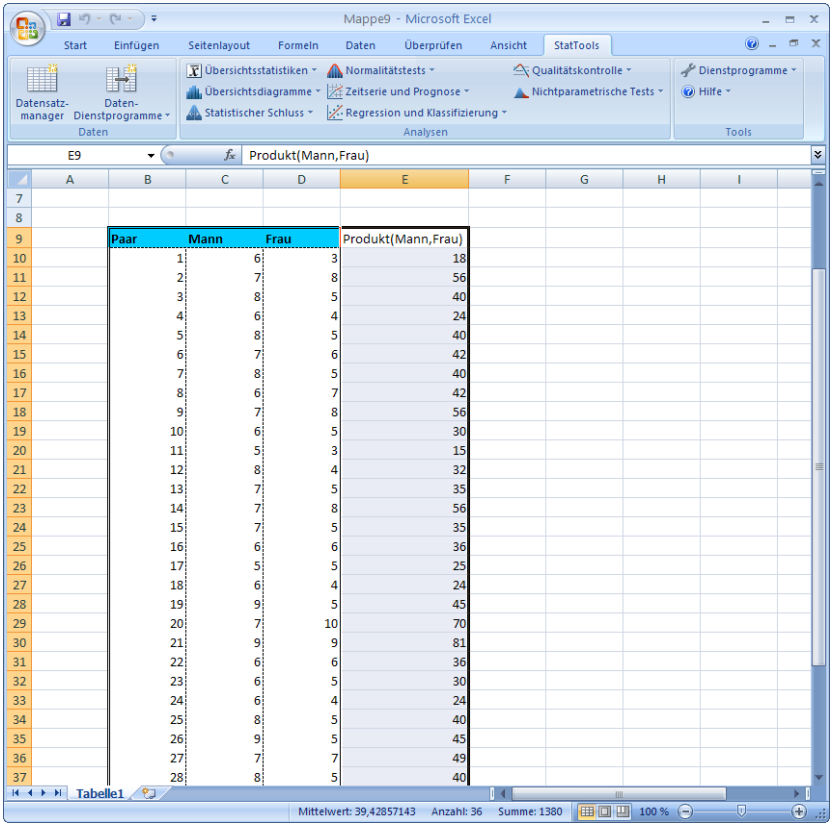


Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Es können jeweils eine oder mehrere Variablen zum Erstellen einer Kombinationsvariable verwendet werden.

In diesem Dialogfeld finden Sie Folgendes:

- **Kombinationstyp** – oder mathematischer Vorgang, der beim Erstellen der Interaktionsvariable an den ausgewählten Variablen ausgeführt werden soll. Es kann sich hierbei um ein Produkt, eine Summe, einen Durchschnitt, ein Minimum, ein Maximum oder um den Min-Max-Bereich handeln.

Datensatz mit Kombinations-variable



Paar	Mann	Frau	Produkt(Mann,Frau)
1	6	3	18
2	7	8	56
3	8	5	40
4	6	4	24
5	8	5	40
6	7	6	42
7	8	5	40
8	6	7	42
9	7	8	56
10	6	5	30
11	5	3	15
12	8	4	32
13	7	5	35
14	7	8	56
15	7	5	35
16	6	6	36
17	5	5	25
18	6	4	24
19	9	5	45
20	7	10	70
21	9	9	81
22	6	6	36
23	6	5	30
24	6	4	24
25	8	5	40
26	9	5	45
27	7	7	49
28	8	5	40

Befehl *Pseudo*

Erstellt Pseudovariablen (0-1) auf Basis von vorhandenen Variablen

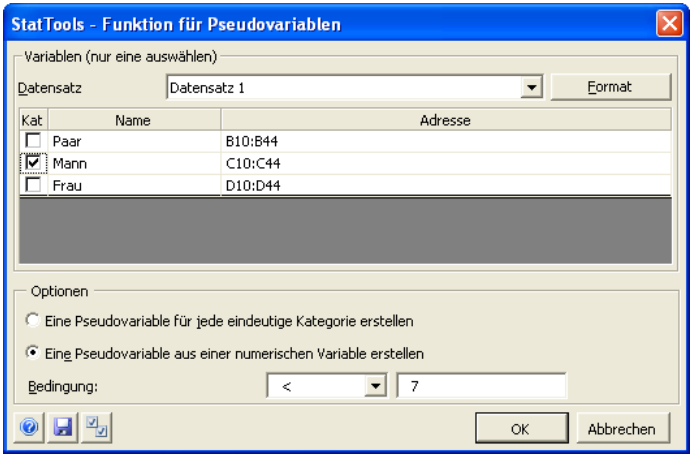
Mithilfe dieses Befehls können auf Basis von vorhandenen Variablen entsprechende Pseudovariablen (0-1) erstellt werden. Zu diesem Zweck sind zwei Optionen verfügbar:

1) Sie können eine Pseudovariable für jede Kategorie einer Kategorievariable erstellen. Wenn Sie z.B. die Kategorievariable *Herkunft* (für Kraftfahrzeuge) mit den Kategorien *USA, Europa* und *Asien* haben, können auf diese Weise Pseudovariablen mit den Namen *Herkunft_USA, Herkunft_Europa* und *Herkunft_Asien* erstellt werden.

2) Sie können aber auch eine Pseudovariable aus einer numerischen Variable erstellen, und zwar auf Basis eines von Ihnen ausgewählten Trennwertes. Wenn Sie beispielsweise mit der Variable *Wertigkeit* arbeiten, könnten Sie Pseudovariablen für die Bedingung *Wertigkeit* <= *160* erstellen. In diesem Fall würde eine neue Variable hinzugefügt werden, die bei *Wertigkeit* > *160* den Wert 0 und bei *Wertigkeit* <= *160* den Wert 1 zeigt.

***Dialogfeld* Funktion für Pseudovariablen**

Zum Erstellen von Pseudovariablen wird das Dialogfeld **Funktion für Pseudovariablen** verwendet:

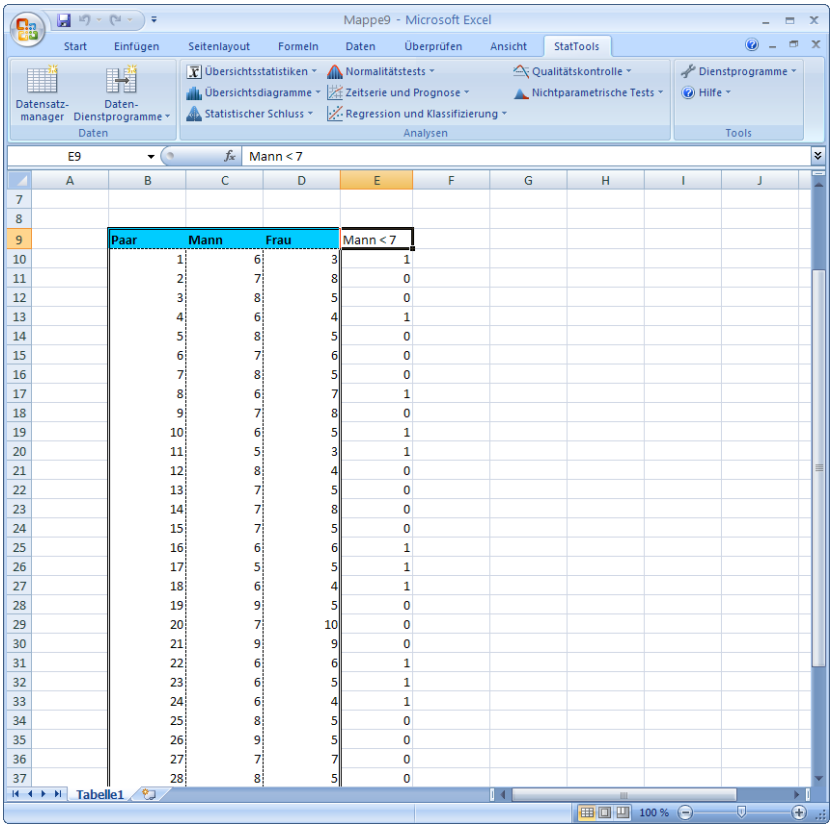


Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Zum Erstellen von Pseudovariablen kann nur jeweils eine Variable verwendet werden.

In diesem Dialogfeld finden Sie Folgendes:

- **Optionen** – legt fest, auf welche Weise die Pseudovariablen erstellt werden sollen – 1) eine Pseudovariable für jede Kategorie einer Kategorievariable oder 2) eine Pseudovariable aus einer numerischen Variable. Durch den **Trennwert** wird die Abgrenzung angegeben, wenn eine numerische Variable einer 0-1-Pseudovariable zugewiesen wird.

Datensatz mit Pseudovariable



Paar	Mann	Frau	Mann < 7
1	6	3	1
2	7	8	0
3	8	5	0
4	6	4	1
5	8	5	0
6	7	6	0
7	8	5	0
8	6	7	1
9	7	8	0
10	6	5	1
11	5	3	1
12	8	4	0
13	7	5	0
14	7	8	0
15	7	5	0
16	6	6	1
17	5	5	1
18	6	4	1
19	9	5	0
20	7	10	0
21	9	9	0
22	6	6	1
23	6	5	1
24	6	4	1
25	8	5	0
26	9	5	0
27	7	7	0
28	8	5	0

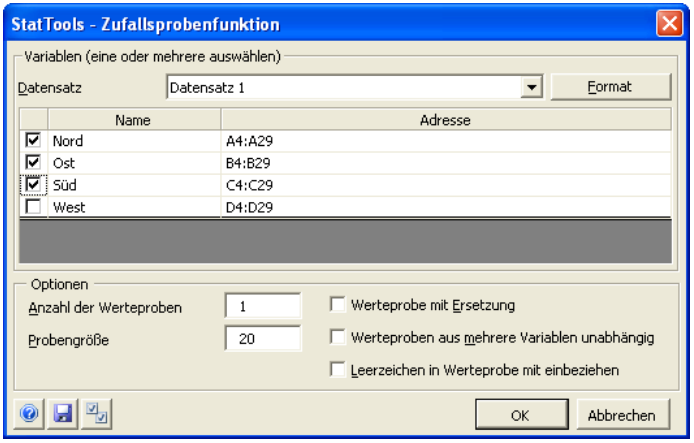
Befehl *Zufallsprobe*

Erstellt eine beliebige Anzahl von Zufallswerteproben aus den ausgewählten Variablen

Dieser Befehl ermöglicht Ihnen, Zufallsproben aus den von Ihnen ausgewählten Variablen zu generieren. Sie müssen die Anzahl der Werteproben und die Größe jeder Probe angeben und durch StatTools werden dann die entsprechenden Werteproben aus den ausgewählten Variablen generiert. Proben können unabhängig oder auch abhängig voneinander aus mehreren Variablen erhoben werden und die Probenerhebung kann entweder mit oder auch ohne Ersetzung vorgenommen werden.

***Dialogfeld* Zufallsproben-funktion**

Zum Generieren von Zufallsproben wird das Dialogfeld **Zufallsprobenfunktion** verwendet:

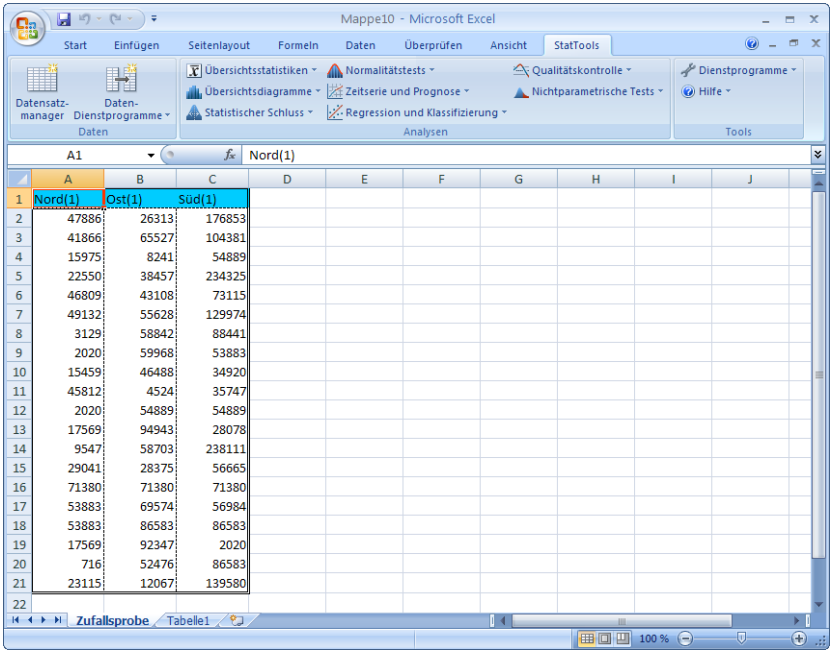


Bei dem ausgewählten Datensatz handelt es sich immer um entstapelte Daten. Es können jeweils eine oder mehrere Variablen zum Generieren von Zufallsproben verwendet werden.

Unter **Optionen** finden Sie In diesem Dialogfeld Folgendes:

- **Anzahl der Werteproben** und **Probengröße –** Für jede ausgewählte Variable wird die angegebene Anzahl an Werteproben generiert und jede Werteprobe enthält eine Anzahl von Elementen, die genau der Probengröße entspricht.
- **Werteprobe mit Ersetzung –** weist darauf hin, dass die Werteprobe nach der Probenerhebung in der Originalpopulation ersetzt wird, sodass diese Werteprobe erneut erhoben werden kann. Wenn **Werteprobe mit Ersetzung** dagegen nicht aktiviert ist (d.h., wenn **Werteprobe ohne Ersetzung** verwendet wird), wird die Werteprobe in der Originalpopulation nicht ersetzt und kann daher auch nicht erneut erhoben werden.
- **Werteproben aus mehrere Variablen unabhängig –** legt fest, dass eine unabhängige Probenerhebung für jede Werteprobe in jeder Variable ausgeführt werden soll. Andernfalls wird in allen Variablen für jede Werteprobe derselbe Probenerhebungsindex (d.h., eine Zahl zwischen 1 und der Werteanzahl in der Variable) verwendet.

Erstellte Zufallsproben



	A	B	C
1	Nord(1)	Ost(1)	Süd(1)
2	47886	26313	176853
3	41866	65527	104381
4	15975	8241	54889
5	22550	38457	234325
6	46809	43108	73115
7	49132	55628	129974
8	3129	58842	88441
9	2020	59968	53883
10	15459	46488	34920
11	45812	4524	35747
12	2020	54889	54889
13	17569	94943	28078
14	9547	58703	238111
15	29041	28375	56665
16	71380	71380	71380
17	53883	69574	56984
18	53883	86583	86583
19	17569	92347	2020
20	716	52476	86583
21	23115	12067	139580

Menü *Übersichtsstatistik*

Mithilfe der Befehle im Menü **Übersichtsstatistik** können Sie mehrere numerische Übersichtsfaktoren für einzelne Variablen oder Variablenpaare berechnen. Hinweis: In den StatTools-Vorgängen gibt es keine Eventualfalltabellen. Anstelle solcher Tabellen werden die Pivot-Tabellen in Excel verwendet.

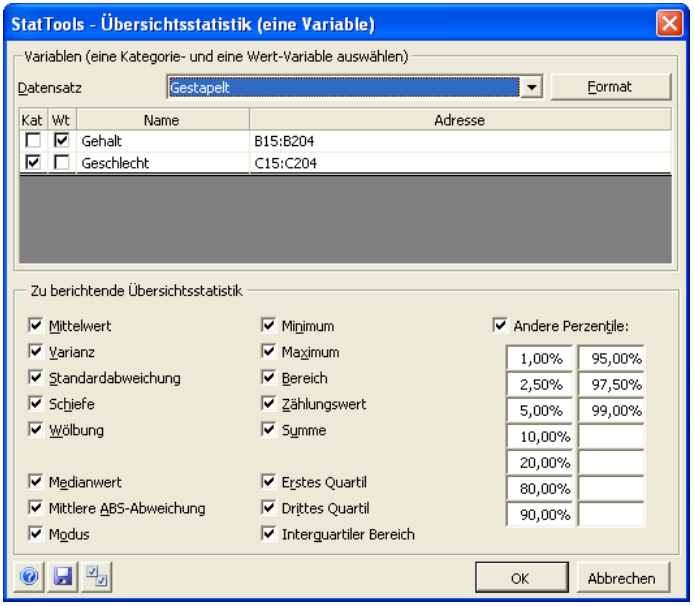
Befehl *Übersicht (eine Variable)*

Berechnet die Übersichtsstatistik für Variablen

Über diesen Befehl können Übersichtsdaten für eine beliebige Anzahl von ausgewählten numerischen Variablen angezeigt werden. Das bezieht Variablen für Mittelwert, Medianwert, Standardabweichung, Varianz, Minimum, Maximum, Bereich, mittlere absolute Abweichung, Schiefe, Wölbung, Zählwert, Summe und ausgewählte Perzentile mit ein.

***Dialogfeld* Übersichtsstatistik (eine Variable)**

Zum Definieren dieser Analyse wird das Dialogfeld **Übersichtsstatistik (eine Variable)** verwendet.



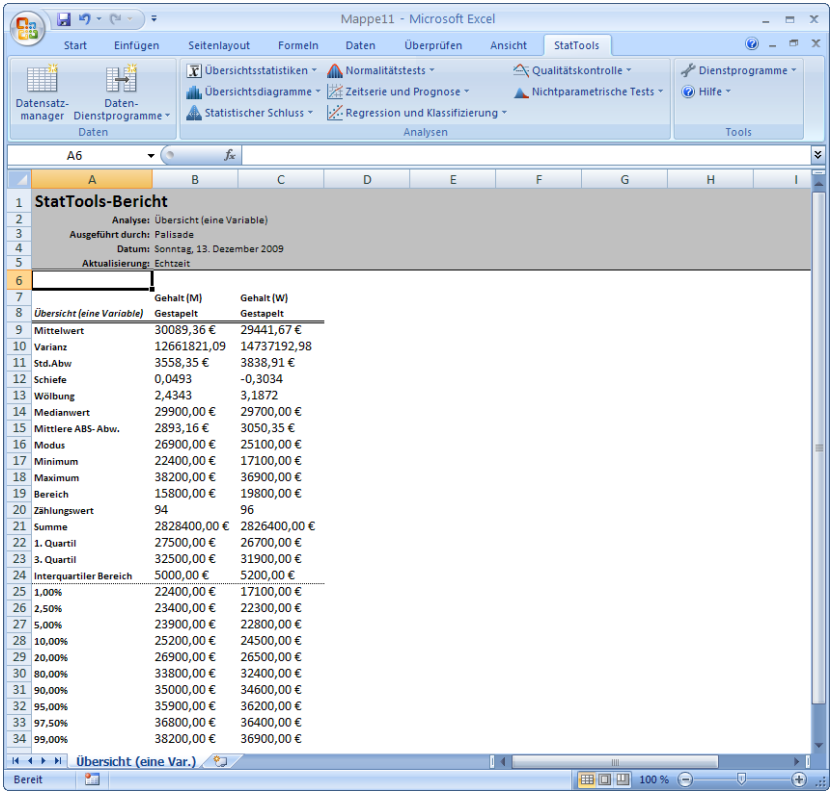
Für diese Analyse können mehrere Variablen ausgewählt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Variablen können aus verschiedenen Datensätzen stammen.

Das Dialogfeld **Übersichtsstatistik (eine Variable)** enthält die Option

- **Zu berichtende Übersichtsstatistik** – um die Statistiken auszuwählen, die in den Bericht mit eingezogen werden sollen. Durch Eingeben der gewünschten Werte können Perzentile hinzugefügt werden.

Übersichtsbericht (eine Variable)

In diesem Bericht werden StatTools-Funktionen (wie z.B. **StatSkewness**) verwendet, um die automatische Datenverknüpfung zu ermöglichen. Der Bericht wird wie unter **Einstellungen** angegeben gespeichert.



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Über diesen Befehl können fehlende Daten fallweise ignoriert werden. Mit anderen Worten, beim Berechnen der Übersichtsfaktoren können für die einzelnen Variablen die fehlenden Daten ignoriert werden. (Das ist sowieso in Excel die Standardmethode. Wird beispielsweise die Funktion MITTELWERT auf einen Bereich angewendet, wird dadurch nur der Durchschnitt der *numerischen* Werte im Bereich genommen.)
- **Mit Daten verknüpfen –** Alle Übersichtsfaktoren werden durch Formeln berechnet, die mit den Daten verknüpft sind. Bei Änderung irgendwelcher Daten ändern sich die Übersichtsfaktoren dann ganz automatisch entsprechend.

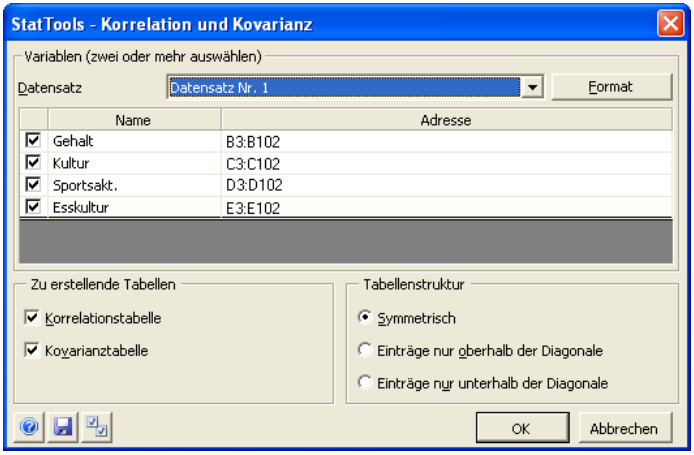
Befehl *Korrelationen und Kovarianz*

Erstellt eine Tabelle der Korrelationen und/oder eine Tabelle der Kovarianzen unter den Variablen

Über diesen Befehl kann eine Tabelle der Korrelationen oder eine Tabelle der Kovarianzen in einem Satz von ausgewählten numerischen Variablen erstellt werden. Da diese beiden Tabellen symmetrisch sind (d.h., die Korrelation zwischen x und y die gleiche wie zwischen y und x ist), können Sie 1) nur die Korrelationen (oder Kovarianzen) *unterhalb* der Diagonale oder 2) nur solche *oberhalb* der Diagonale oder 3) alle Korrelationen *unterhalb* und *oberhalb* der Diagonale anzeigen lassen.

***Dialogfeld* Korrelationen und Kovarianz**

Zum Definieren dieser Analyse wird das Dialogfeld **Korrelationen und Kovarianz** verwendet:



Für diese Analyse müssen zwei oder mehr Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Variablen können aus verschiedenen Datensätzen stammen.

In diesem Dialogfeld sind die Optionen wie folgt unterteilt:

- **Zu erstellende Tabellen** – wählt die gewünschte Korrelations- und/oder Kovarianztabelle aus

- **Tabellenstruktur –** legt die Struktur der zu erstellenden Tabellen fest:

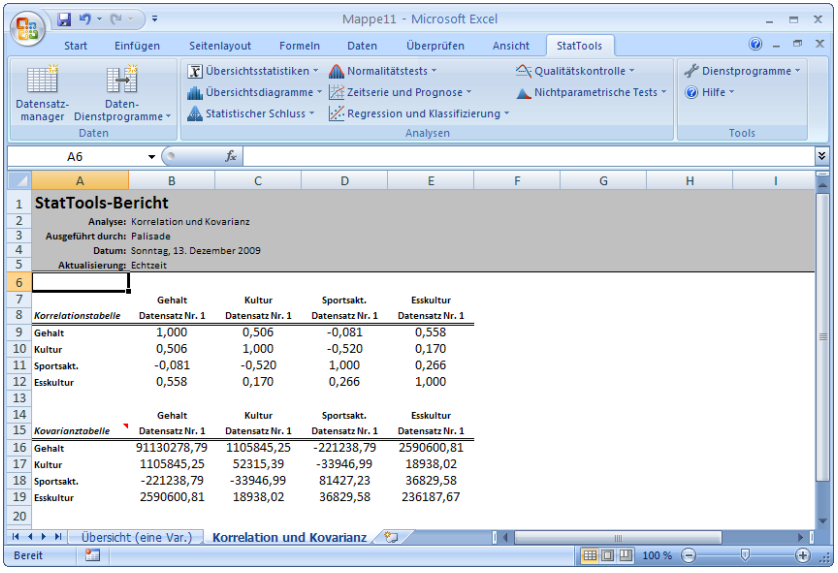
  - **Symmetrisch** – Korrelationen (oder Kovarianzen) werden *unterhalb* und *oberhalb* der Diagonale angezeigt

  - **Einträge nur oberhalb der Diagonale** – Korrelationen (oder Kovarianzen) werden nur *oberhalb* der Diagonale angezeigt.

  - **Einträge nur unterhalb der Diagonale** – Korrelationen (oder Kovarianzen) werden nur *unterhalb* der Diagonale angezeigt.

Bericht über Korrelationen und Kovarianz

In diesem Bericht werden StatTools-Funktionen (wie z.B. **StatCorrelationCoeff**) verwendet, um die automatische Datenverknüpfung zu ermöglichen. Der Bericht wird wie unter **Einstellungen** angegeben gespeichert.



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Bei Variablenpaaren kann mit fehlenden Daten gearbeitet werden. Mit anderen Worten, um die Korrelation (oder Kovarianz) wischen einem Variablenpaar zu erreichen, werden alle Fälle, in denen bei einer der beiden Variablen irgendwelche Daten fehlen, einfach ignoriert.

- **Mit Daten verknüpfen –** Die Korrelationen und Kovarianzen werden durch Formeln berechnet, die mit den Daten verknüpft sind. Bei Änderung irgendwelcher Daten werden diese Übersichtsfaktoren daher ganz automatisch aktualisiert.

Menü *Übersichtsdiagramme*

Über die Befehle im Menü **Übersichtsdiagramme** können Sie Diagramme erstellen, die für die Statistikanalyse sehr nützlich sind und nicht so leicht mittels Diagramm-Assistent in Excel generiert werden können. Natürlich ist uns klar, dass Excel bereits umfangreiche Diagrammerstellungsmöglichkeiten enthält. In StatTools wird daher nicht versucht, bereits vorhandene Fähigkeiten zu duplizieren.

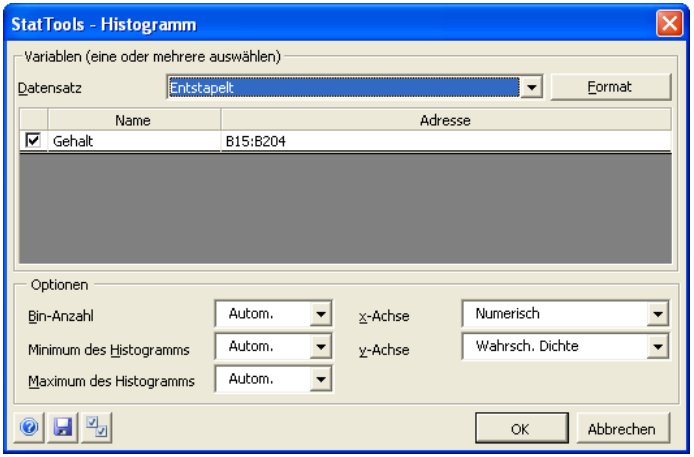
Befehl *Histogramm*

Erstellt Histogramme für Variablen

Über diesen Befehl kann für jede von Ihnen ausgewählte Variable ein Histogramm erstellt werden. Dabei wird Ihnen die Möglichkeit gegeben, die Kategorien (auch „Bins“ genannt) des Histogramms zu definieren und deutlich im Diagramm anzuzeigen. Auch kann auf diese Weise die Häufigkeitstabelle des Histogramms erstellt werden.

***Dialogfeld* Histogramm**

Zum Definieren dieses Diagrammtyps wird das Dialogfeld **Histogramm** verwendet:

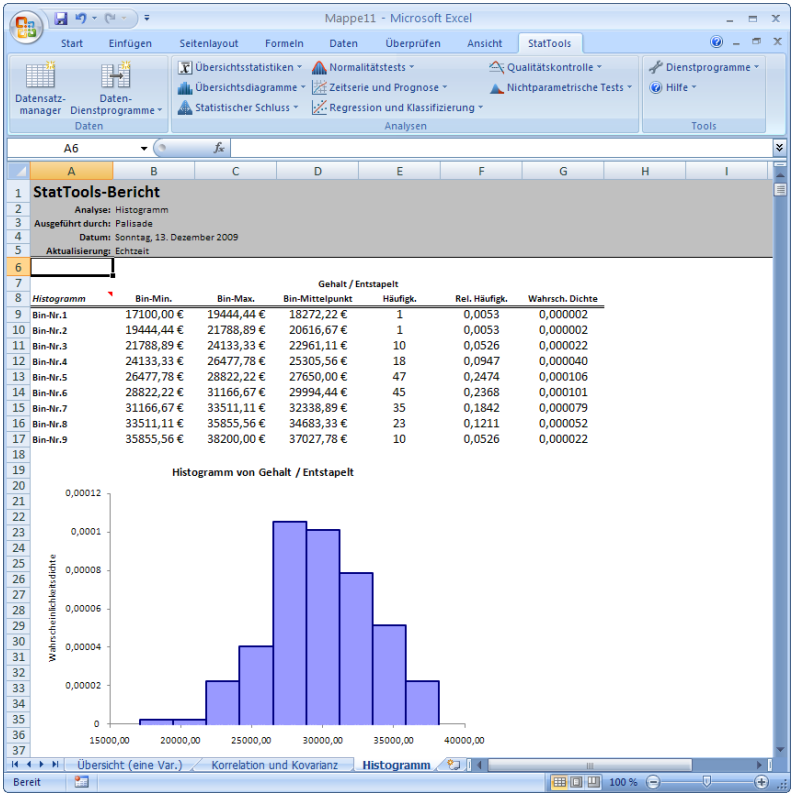


Für die grafische Darstellung können eine oder mehrere Variablen ausgewählt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Diese Variablen können aus verschiedenen Datensätzen stammen.

Unter **Optionen** ist im Dialogfeld **Histogramm** Folgendes zu finden:

- **Bin-Anzahl –** legt die Anzahl der Histogramm-Intervalle fest, die quer über den Diagrammbereich berechnet werden sollen. Der eingegebene Wert muss im Bereich von 1 bis 200 liegen. Über die Einstellung **Automatisch** kann auf Basis einer internen heuristischen Methode die beste Bin-Anzahl für die gegebenen Daten berechnet werden.
- **Minimum des Histogramms –** stellt den Minimalwert für den Beginn der Histogramm-Bins ein. Durch **Automatisch** wird angegeben, dass StatTools die Histogramm-Bins auf Basis des Minimalwerts der grafisch dargestellten Daten beginnen soll.
- **Maximum des Histogramms –** stellt den Maximalwert für das Beenden der Histogramm-Bins ein. Durch **Automatisch** wird angegeben, dass StatTools die Histogramm-Bins auf Basis des Maximalwerts der grafisch dargestellten Daten enden soll.
- **x-Achse –** wählt eine **kategorische** oder **numerische** x-Achse aus. Bei einer kategorischen x-Achse wird jedes Bin einfach durch den Mittelpunkt des Bins gekennzeichnet. Auf einer numerischen x-Achse ist dagegen ein „lesbares" Minimum und Maximum der Achse zu sehen, das mithilfe der standardmäßigen Excel-Neuskalieroptionen anders skaliert werden kann.
- **y-Achse –** wählt **Häufigkeit, Rel. Häufigkeit** oder **Wahrsch. Dichte** als die Messwerteinheit, die auf der y-Achse berichtet wird. Häufigkeit bezeichnet die effektive Anzahl der Beobachtungen in einem Bin. **Relative Häufigkeit** ist die Wahrscheinlichkeit, dass ein bestimmter Wert in einem Bereich von Bins auftritt (Beobachtungen in einem Bin/Beobachtungen insgesamt). **Dichte** ist die relative Häufigkeit, dividiert durch die Bin-Breite. Dadurch wird sichergestellt, dass die Werte der y-Achse bei Änderung der Bin-Anzahl stets konstant bleiben.

Histogramm einer einzelnen Variable



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Histogramme sind teilweise mit Daten verknüpft. Wenn sich Daten ändern und neue Werte in den Bereich der Bins des Originalhistogramms fallen, wird das Diagramm automatisch aktualisiert. Datenänderungen, die ein erneutes Binning erforderlich machen, verursachen jedoch keine Aktualisierung des Diagramms.

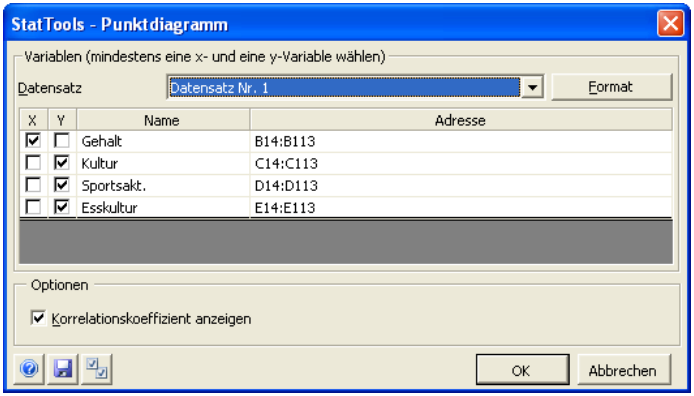
Befehl *Punktdiagramm*

Erstellt Punktdiagramme zwischen Variablenpaaren

Mithilfe dieses Befehls kann ein Punktdiagramm für jedes von Ihnen ausgewählte Variablenpaar erstellt werden. Punktdiagramme können auch mittels *xy*-Diagrammoption in Excel erstellt werden, aber dann wird die erste (oder ganz links befindliche) Variable auf der Horizontalachse platziert, was vielleicht nicht Ihren Wünschen entspricht. In StatTools können Sie dagegen auswählen, welche der Variablen auf der horizontalen Achse platziert werden soll. In jedem Punktdiagramm ist die Korrelation zwischen den beiden Variablen und dem entsprechenden Diagramm zu sehen.

***Dialogfeld* Punktdiagramm**

Zum Definieren dieses Diagrammtyps wird das Dialogfeld **Punktdiagramm** verwendet:



Für die grafische Darstellung können zwei oder mehr Variablen ausgewählt werden. Es ist mindestens eine *x*-Achsen- und eine *y*-Achsen-Variable erforderlich. Falls mehr als zwei Variablen ausgewählt sind, werden mehrere Punktdiagramme erstellt. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Diese Variablen können aus verschiedenen Datensätzen stammen.

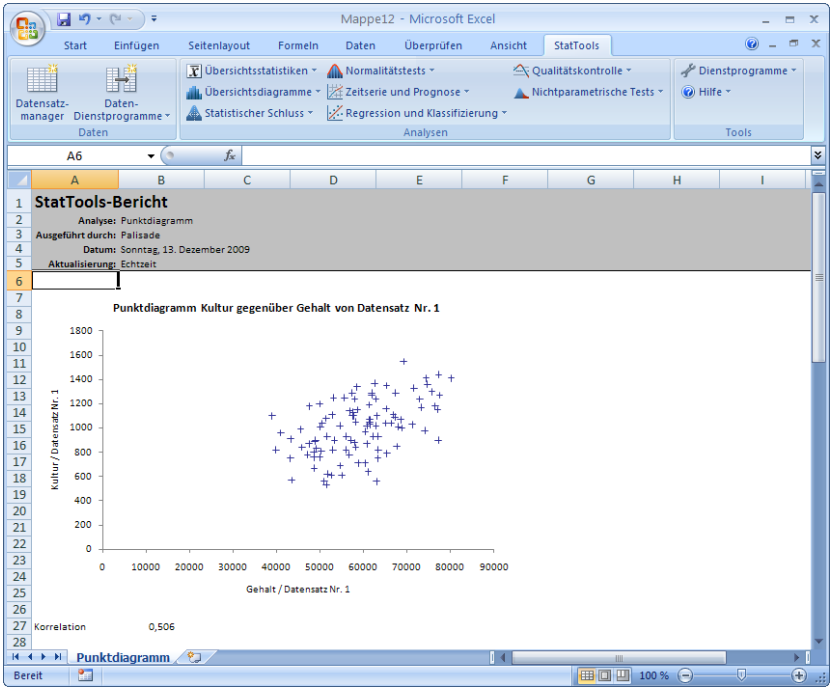
Unter **Optionen** ist im Dialogfeld **Punktdiagramm** Folgendes zu finden:

- **Korrelationskoeffizient anzeigen** – legt fest, dass der Korrelationskoeffizient zwischen den grafisch dargestellten Variablen angezeigt werden soll.
- **Diagrammtyp.** Ermöglicht die Auswahl des zu erstellenden Diagrammtyps. Es sind folgende Optionen verfügbar:

  Einfaches xy-Diagramm. zum Erstellen eines typischen Excel-Punktdiagramms.

  Unterteilt nach Kategorienvariable. Die Punkte im Punktdiagramm sind gefärbt, und zwar auf Basis der Kategorien, zu denen die Datenpunkte gehören (siehe zusätzliche Kategorien-Variable).

Punktdiagramm (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In jedem ausgewählten Variablenpaar werden Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Die Punktdiagramme sind mit den Originaldaten verknüpft. Sobald sich die Daten ändern, werden die Punktdiagramme entsprechend aktualisiert. Die Achsenskalen müssen jedoch u.U. manuell aktualisiert werden, falls sich die Bereiche der ausgewählten Variablen erheblich ändern.

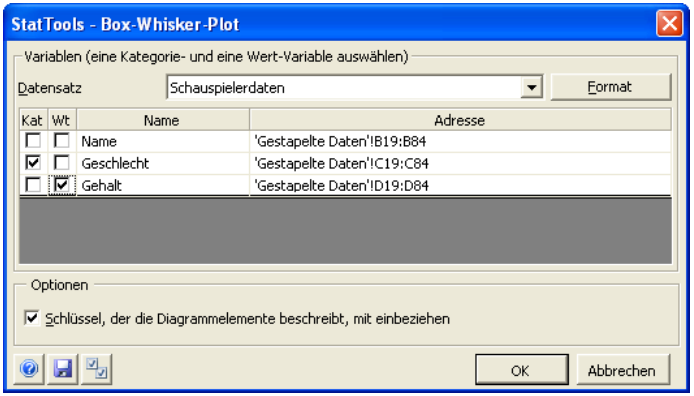
Befehl *Box-Whisker-Plot*

Erstellt Box-Whisker-Plots für Variablen

Über diesen Befehl können Sie bei Auswahl einer Variable ein einzelnes Box-Whisker-Plot oder bei Auswahl mehrerer Variablen mehrere Box-Whisker-Plots nebeneinander erstellen. Auch wird dadurch ein Blatt erstellt, auf dem die Übersichtsstatistik (d.h., Quartile, interquartile Bereiche usw.) zu sehen ist, die zum Definieren des Box-Whisker-Plots verwendet wird.

***Dialogfeld* Box-Whisker-Plot**

Zum Definieren dieses Diagrammtyps wird das Dialogfeld **Box-Whisker-Plot** verwendet:

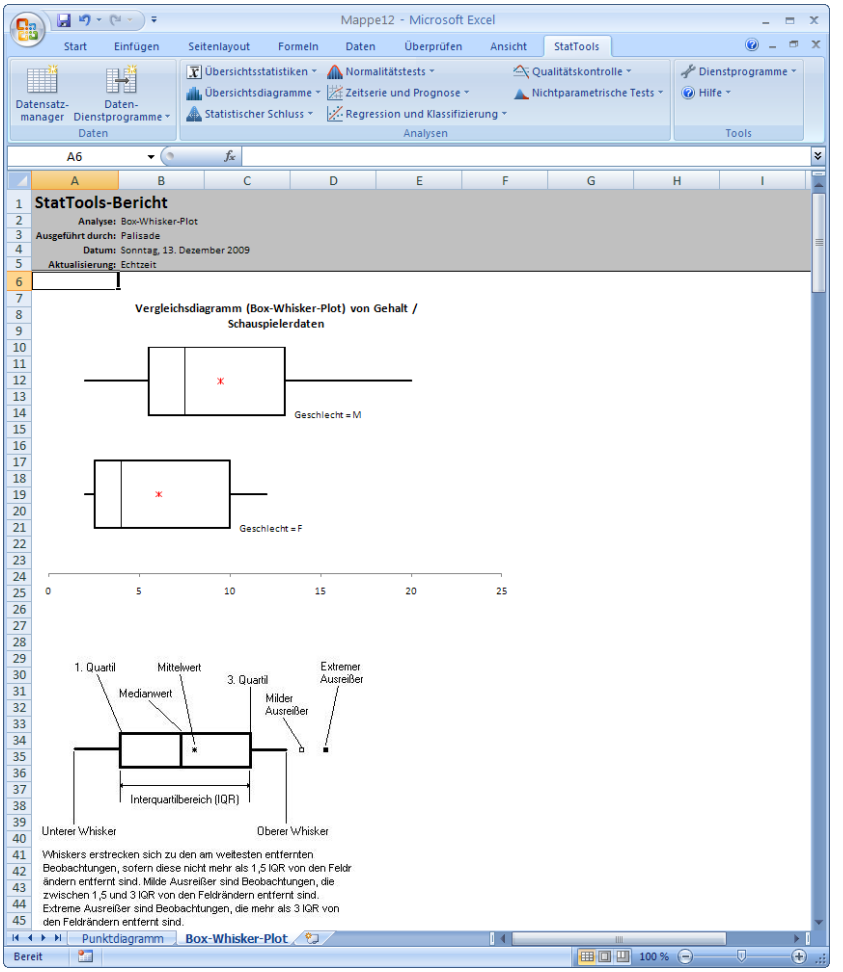


Für die grafische Darstellung können eine oder mehrere Variablen ausgewählt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Diese Variablen können aus verschiedenen Datensätzen stammen.

Unter **Optionen** ist im Dialogfeld **Box-Whisker-Plot** Folgendes zu finden:

- **Schlüssel, der die Diagrammelemente beschreibt, mit einbeziehen** – legt fest, dass ein separater Schlüssel, der die Diagrammelemente beschreibt, unterhalb des Diagramms angezeigt werden soll.

Box-Whisker-Plot (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den für das Diagramm ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Die erstellten Box-Whisker-Plots sind mit den Originaldaten verknüpft, d.h., sobald sich die Daten ändern, werden die Diagramme automatisch entsprechend aktualisiert. Die Skala der horizontalen Achse muss jedoch u.U. manuell aktualisiert werden, falls sich die Daten erheblich ändern.

Menü *Statistischer Schluss*

Mithilfe der Befehle in diesem Menü können die gebräuchlichsten Statistischer Schluss-Analysen ausgeführt werden, d.h., Vertrauensintervall- und Hypothesetests zusammen mit einseitiger und zweiseitiger ANOVA.

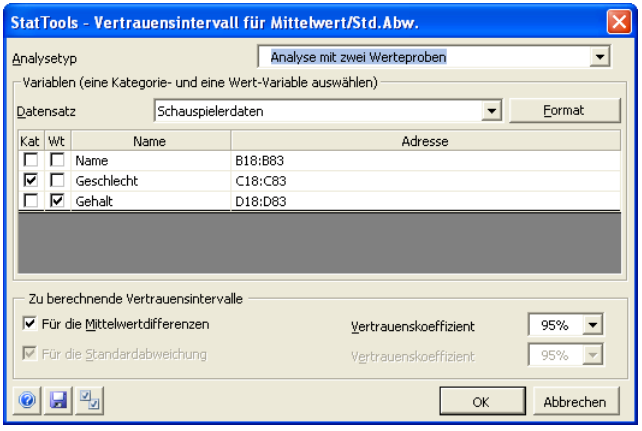
Befehl *Vertrauensintervall – Mittelwert/Std. Abweichung*

Berechnet die Vertrauensintervalle für Mittelwert und Standardabweichung von Variablen

Über diesen Befehl können Sie das Vertrauensintervall für Mittelwert und Standardabweichung von einzelnen Variablen berechnen. Auch kann auf diese Weise die Differenz zwischen den Mittelwerten berechnet werden, wenn es sich um Variablenpaare handelt. Die Vertrauensintervalle können mithilfe einer Analyse berechnet werden, und zwar unter Verwendung einer, zweier oder gepaarter Werteproben.

Zum Definieren dieser Analyse wird das Dialogfeld **Vertrauensintervall für Mittelwert/Std.Abw.** verwendet:

Dialogfeld Vertrauensintervall für Mittelwert/Std.Abw.

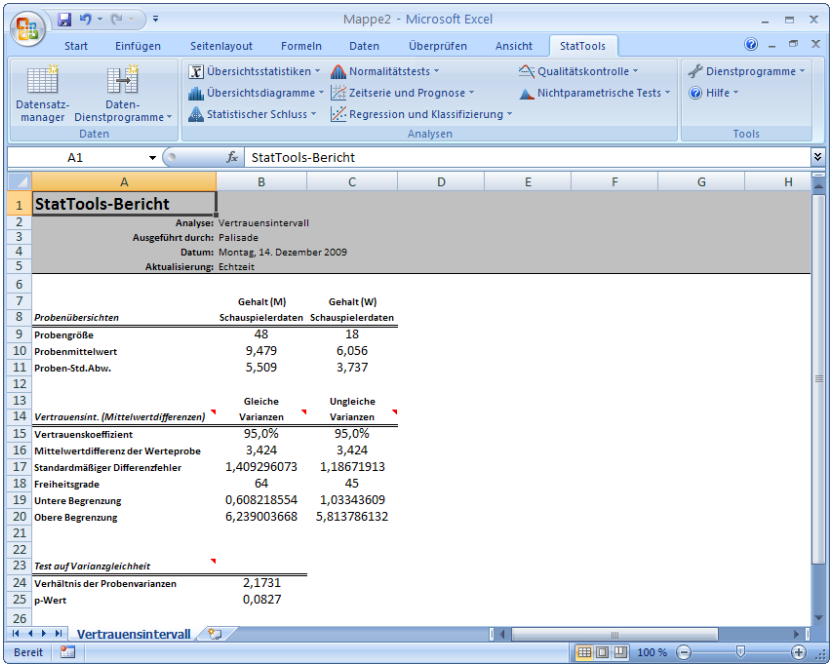


Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Eine Analyse unter Verwendung nur einer Werteprobe erfordert eine oder mehrere Variablen, während für eine Analyse mit zwei Werteproben oder mit einer gepaarten Werteprobe zwei Variablen benötigt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Diese Variablen können aus verschiedenen Datensätzen stammen.

Im Dialogfeld **Vertrauensintervall für Mittelwert/Std.Abw.** sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Hierfür gibt es folgende Möglichkeiten:
  - **Analyse mit einer Werteprobe –** berechnet die Vertrauensintervalle für einen einzelne numerische Variable
  - **Analyse mit zwei Werteproben –** berechnet das Vertrauensintervall für die Differenz zwischen den Mittelwerten aus zwei unabhängigen Populationen
  - **Analyse mit gepaarter Werteprobe –** Diese Analyse entspricht praktisch der Analyse mit zwei Werteproben, wird aber bevorzugt, wenn die beiden Variablen irgendwie auf natürliche Weise gepaart sind. Im Wesentlichen wird hier eine Analyse mit einer Werteprobe in Bezug auf die Differenzen zwischen Variablenpaaren ausgeführt.
- **Zu berechnende Vertrauensintervalle –** legt die Vertrauensintervalle fest, die für die ausgewählten Variablen berechnet werden sollen. Die Optionen sind je nach ausgewähltem Analysetyp unterschiedlich:
  - **Analyse mit einer Werteprobe –** wird ausgewählt, um die Vertrauensintervalle bezüglich Mittelwert und/oder Standardabweichung zu berechnen sowie auch den Vertrauenskoeffizienten (zwischen 0 und 100 %) für Mittelwert und Standardabweichung.
  - **Analyse mit zwei Werteproben oder gepaarter Werteprobe –** wird gewählt, um das Vertrauensintervall für die Differenz zwischen den Mittelwerten von zwei Variablen zu berechnen und um den gewünschten Vertrauenskoeffizienten (zwischen 0 und 100 %) festzulegen.

Bericht über Vertrauensintervall



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

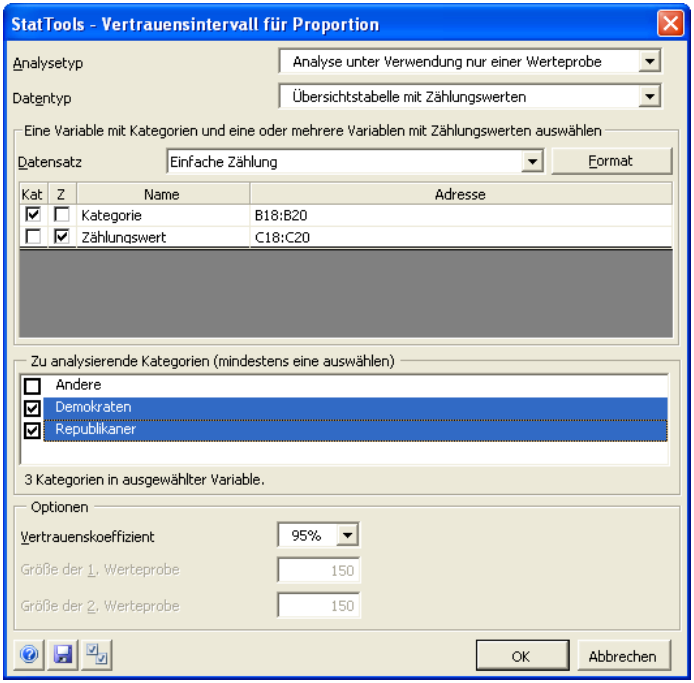
Befehl *Vertrauensintervall für Proportionen*

Berechnet die Vertrauensintervalle für Proportionen

Sie können diesen Befehl zum Analysieren der Proportion von Elementen in einer Werteprobe verwenden, die zu einer bestimmten Kategorie gehört (**Analyse mit einer Werteprobe**) oder zum Vergleichen von zwei Werteproben hinsichtlich der Proportion von Elementen in einer bestimmten Kategorie (**Analyse mit zwei Werteproben**). Durch diesen Vorgang werden drei verschiedene Datentypen unterstützt: **Populationsprobe, Übersichtstabelle mit Zählungswerten** und **Übersichtstabelle mit Proportionen**.

Zum Definieren dieser Analyse wird das Dialogfeld **Vertrauensintervall für Proportion** verwendet:

Dialogfeld Vertrauensintervall für Proportion



Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Für eine **Analyse mit einer Werteprobe** sind eine oder mehrere Variablen erforderlich, während für die **Analyse mit zwei Werteproben** zwei Variablen benötigt werden. Zum Auswählen von Variablen mit Werteprobeninformationen wird die Spalte **Ct** (Zählwert), % (Proportion) oder die Spalte ohne Bezeichnung für den Datentyp **Populationsprobe** verwendet. (Wenn die Daten der Populationsprobe gestapelt sind, werden die Werteproben in den Spalten **C1** und **C2** ausgewählt, wobei **C1** die gestapelten Kategorien enthält.) Wenn die Daten dagegen in Form einer Tabelle mit Zählwerten und Proportionen auftreten, ist die zusätzliche Spalte **Kat.** (Kategorie) vorhanden, um eine Variable mit Kategorienamen auszuwählen.

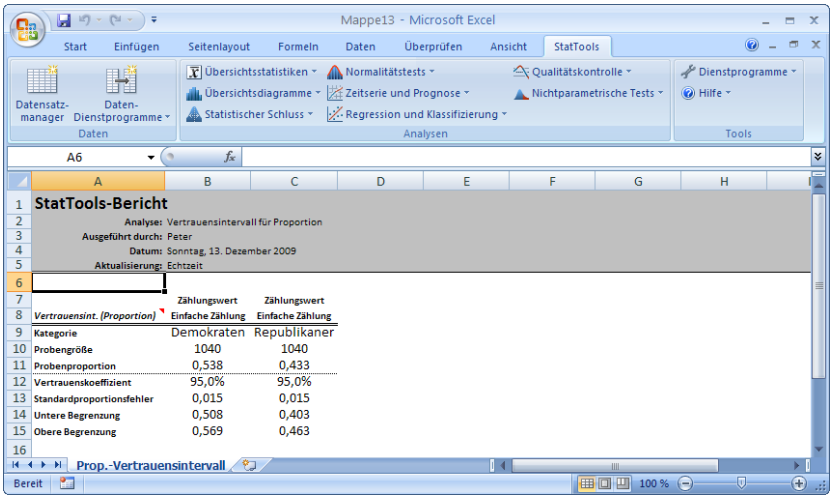
Bei dem Datentyp **Populationsprobe** können die Variablen aus verschiedenen Datensätzen stammen..

Im Dialogfeld **Vertrauensintervall für Proportion** sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Hierfür gibt es folgende Möglichkeiten:
  - **Analyse mit einer Werteprobe –** Berechnet die Vertrauensintervalle für die Proportion von Elementen in einer Werteprobe, die zu einer bestimmten Kategorie gehört.
  - **Analyse mit zwei Werteproben –** berechnet das Vertrauensintervall für zwei Werteproben mit Bezug auf die Proportion von Elementen in einer bestimmten Kategorie.
- **Datentyp –** legt den Typ der zu analysierenden Daten fest, entweder **Populationsprobe, Übersichtstabelle mit Zählungswerten** oder **Übersichtstabelle mit Proportionen**.
- **Optionen –** Die jeweils verfügbaren Optionen hängen vom Analysetyp und vom Datentyp ab. Es kann sich dabei um folgende Optionen handeln:
  - **Vertrauenskoeffizient –** wählt den Vertrauenskoeffizienten (zwischen 0 und 100 %) für die Analyse aus.
  - **Größe der 1. und 2. Werteprobe –** wählt für den Datentyp **Übersichtstabelle mit Proportionen** die Größe

der 1. und 2. Werteprobe aus (nur bei Analyse mit zwei Werteproben).

Bericht über Vertrauensintervall



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

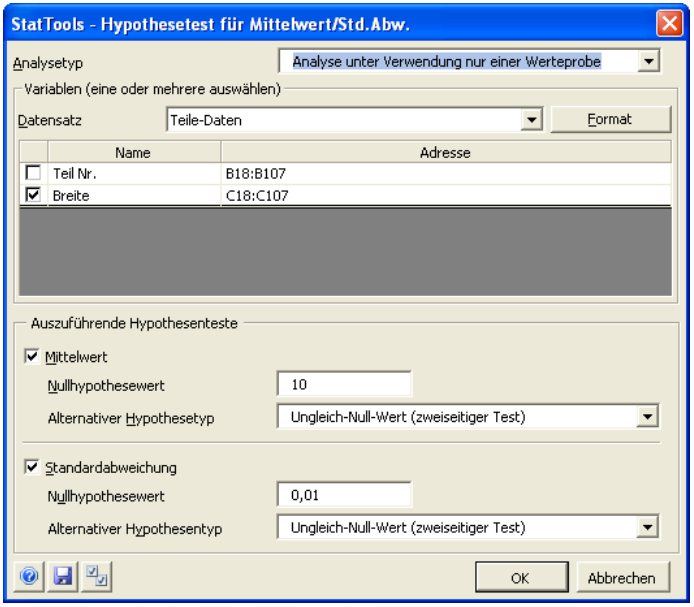
Befehl *Hypothesetest für Mittelwert/Std. Abweichung*

Führt einen Hypothesetest für Mittelwert und Standardabweichung von Variablen aus

Mithilfe dieses Befehls können bei einzelnen Variablen Hypothesetests für Mittelwert und Standardabweichung ausgeführt werden oder auch für die Differenzen zwischen den Mittelwerten, wenn es sich um Variablenpaare handelt. Die Hypotheseteste können mithilfe einer Analyse ausgeführt werden, und zwar unter Verwendung von **einer, zwei** oder **gepaarten Werteproben.**

***Dialogfeld* Hypothesetest für Mittelwert/Std.Abw.**

Zum Definieren dieser Analyse wird das Dialogfeld **Hypothesetest für Mittelwert/Std.Abw.** verwendet:

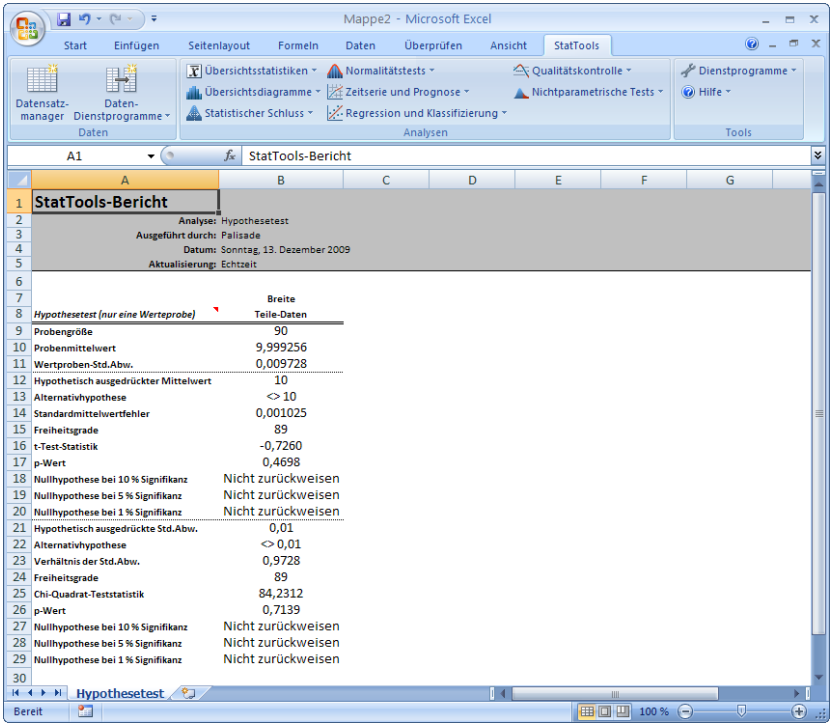


Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Eine Analyse unter Verwendung nur einer Werteprobe erfordert eine oder mehrere Variablen, während für eine Analyse mit zwei Werteproben oder mit einer gepaarten Werteprobe zwei Variablen benötigt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen. Diese Variablen können aus verschiedenen Datensätzen stammen.

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Es kann sich dabei um folgende Optionen handeln:
  - **Analyse mit einer Werteprobe –** führt Hypothesetests für eine einzelne numerische Variable aus.
  - **Analyse mit zwei Werteproben –** führt Hypothesetests in Bezug auf die Differenz zwischen Mittelwerten aus zwei unabhängigen Populationen aus.
  - **Analyse mit gepaarter Werteprobe –** Diese Analyse entspricht praktisch der Analyse mit zwei Werteproben, wird aber bevorzugt, wenn die beiden Variablen irgendwie auf natürliche Weise gepaart sind. Im Wesentlichen wird hier eine Analyse mit einer Werteprobe in Bezug auf die Differenzen zwischen Variablenpaaren ausgeführt.
- **Auszuführende Hypothesenteste –** legt die Hypothesetests fest, die in Bezug auf die ausgewählten Variablen ausgeführt werden sollen. Die Optionen sind je nach ausgewähltem Analysetyp unterschiedlich. Durch **Analyse mit einer Werteprobe** werden Hypothesetests in Bezug auf den Mittelwert und/oder die Standardabweichung ausgewählt. Durch **Analyse mit zwei Werteproben** oder **Analyse mit gepaarter Werteprobe** werden Hypothesetests in Bezug auf Differenz zwischen den Mittelwerten für zwei Variablen ausgewählt. Für jeden ausgewählten Hypothesetest sind folgende Optionen verfügbar:
  - **Nullhypothesenwert –** d.h., der Wert des Populationsparameters unter der Nullhypothese.
  - **Alternativer Hypothesentyp –** d.h., die Alternative zum Nullhypothesenwert, der während der Analyse ausgewertet wird. Der alternative Hypothesetyp kann entweder **einseitig** (d.h. größer oder kleiner als die Nullhypothese) oder **zweiseitig** (d.h., ungleich der Nullhypothese) sein.

Bericht über Hypothesetests



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

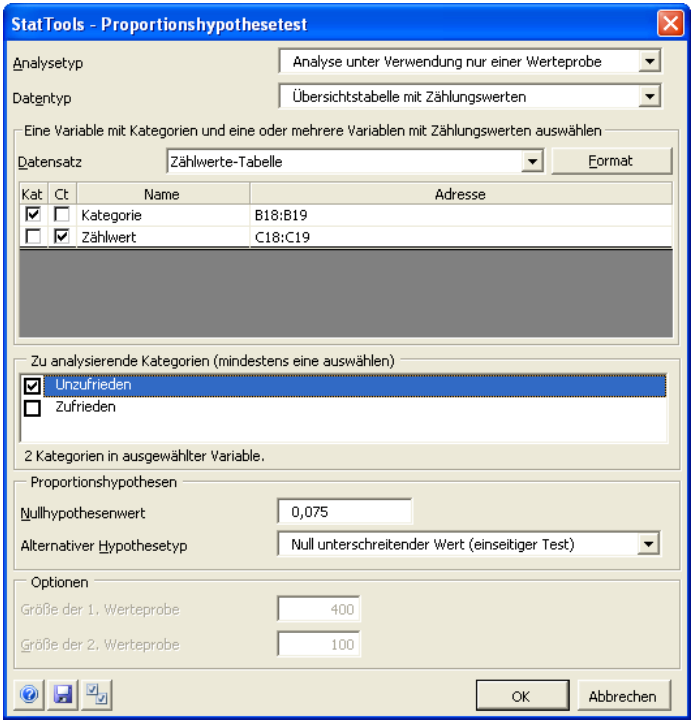
Befehl *Hypothesetest für Proportion*

Führt einen Hypothesetest für Proportionen aus

Über diesen Befehl kann die Proportion von Elementen in einer Werteprobe, die zu einer bestimmten Kategorie gehört (**Analyse mit einer Werteprobe**) analysiert oder können zwei Werteproben hinsichtlich der Proportion von Elementen in einer bestimmten Kategorie (**Analyse mit zwei Werteproben**) verglichen werden. Durch diesen Vorgang werden drei verschiedene Datentypen unterstützt: **Populationsprobe, Übersichtstabelle mit Zählungswerten** und **Übersichtstabelle mit Proportionen**.

Zum Definieren dieser Analyse wird das Dialogfeld **Proportionshypothesetest** verwendet:

Dialogfeld Proportions-hypothesetest



Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Für eine **Analyse mit einer Werteprobe** sind eine oder mehrere Variablen erforderlich, während für die **Analyse mit zwei Werteproben** zwei Variablen benötigt werden. Zum Auswählen von Variablen mit Werteprobeninformationen wird die Spalte **Ct** (Zählwert), % (Proportion) oder die Spalte ohne Bezeichnung für den Datentyp **Populationsprobe** verwendet. (Wenn die Daten der Populationsprobe gestapelt sind, werden die Werteproben in den Spalten **C1** und **C2** ausgewählt, wobei **C1** die gestapelten Kategorien enthält.) Wenn die Daten dagegen in Form einer Tabelle mit Zählwerten und Proportionen auftreten, ist die zusätzliche Spalte **Kat.** (Kategorie) vorhanden, um eine Variable mit Kategorienamen auszuwählen.

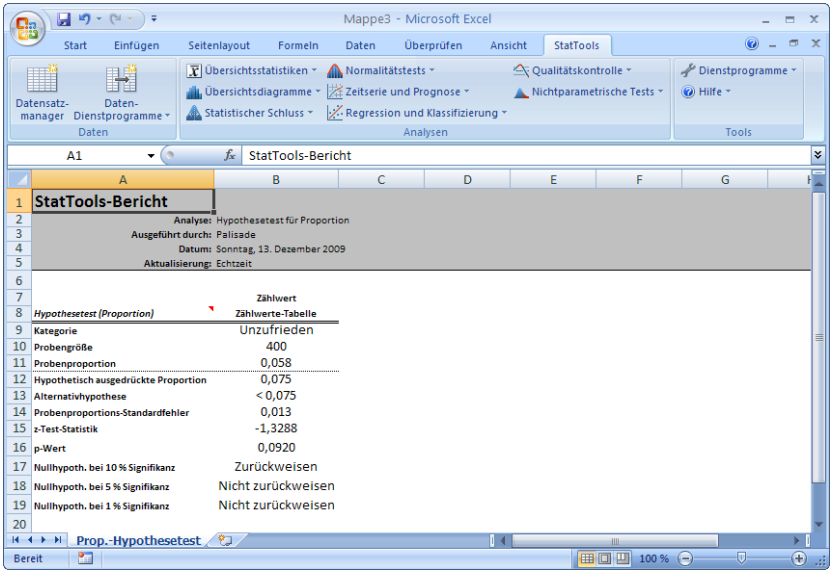
Bei dem Datentyp **Populationsprobe** können die Variablen aus verschiedenen Datensätzen stammen..

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Es kann sich dabei um folgende Optionen handeln:
  - **Analyse mit einer Werteprobe –** führt einen Hypothesetest für die Proportion von Elementen in einer Werteprobe aus, die zu einer bestimmten Kategorie gehört.
  - **Analyse mit zwei Werteproben –** führt einen Hypothesetest für zwei Werteproben mit Bezug auf die Proportion von Elementen in einer bestimmten Kategorie aus.
- **Datentyp –** legt den Typ der zu analysierenden Daten fest, entweder **Populationsprobe, Übersichtstabelle mit Zählungswerten** oder **Übersichtstabelle mit Proportionen**.

- **Proportionshypothesen –** legt die Hypothesetests fest, die in Bezug auf die ausgewählte Proportion ausgeführt werden sollen. Die Optionen sind je nach ausgewähltem Analysetyp unterschiedlich. Es kann sich dabei um folgende Optionen handeln:
  - **Nullhypothesenwert –** d.h., der Wert des Populationsparameters unter Nullhypothese.
  - **Alternativer Hypothesetyp –** d.h., die Alternative zum Nullhypothesenwert, der während der Analyse ausgewertet wird. Der alternative Hypothesetyp kann entweder **einseitig** (d.h. größer oder kleiner als die Nullhypothese) oder **zweiseitig** (d.h., ungleich der Nullhypothese) sein.
- **Größe der 1. und 2. Werteprobe –** wählt für den Datentyp **Übersichtstabelle mit Proportionen** die Größe der 1. und 2. Werteprobe aus (nur bei Analyse mit zwei Werteproben).

Bericht über Hypothesetests



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

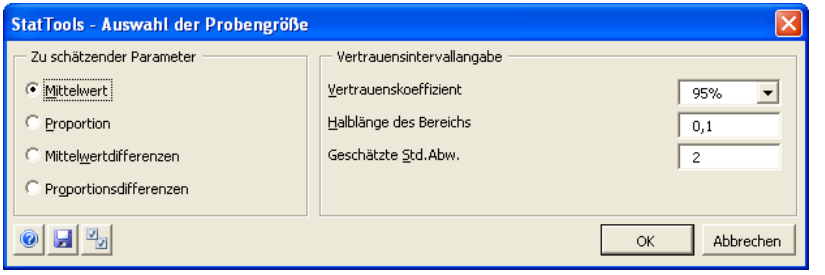
Befehl *Probegrößenauswahl*

Bestimmt die Probengröße, die für Berechnung von Vertrauensintervallen erforderlich ist

Dieser Befehl ermöglicht Ihnen, die Werteprobengröße (oder Probengrößen) zu bestimmen, die erforderlich sind, um ein Vertrauensintervall mit einer vorgeschriebenen Halblänge zu erhalten. Dadurch werden die Vertrauensintervalle für Mittelwert, Proportion, Differenz zwischen Mittelwerten und Differenz zwischen zwei Proportionen festgelegt. Zu diesem Zweck sind keine Datensätze oder Variablen erforderlich, da es sich bei der Probengröße um eine Information handelt, die in der Regel schon *vor* Datenerfassung benötigt wird. Sie müssen jedoch den Vertrauenskoeffizienten, die gewünschte Halblänge und auch andere Parameter angeben, die zum Festlegen der Probengröße erforderlich sind.

***Dialogfeld* Auswahl der Probengröße**

Zum Definieren dieser Analyse wird das Dialogfeld **Auswahl der Probengröße** verwendet:

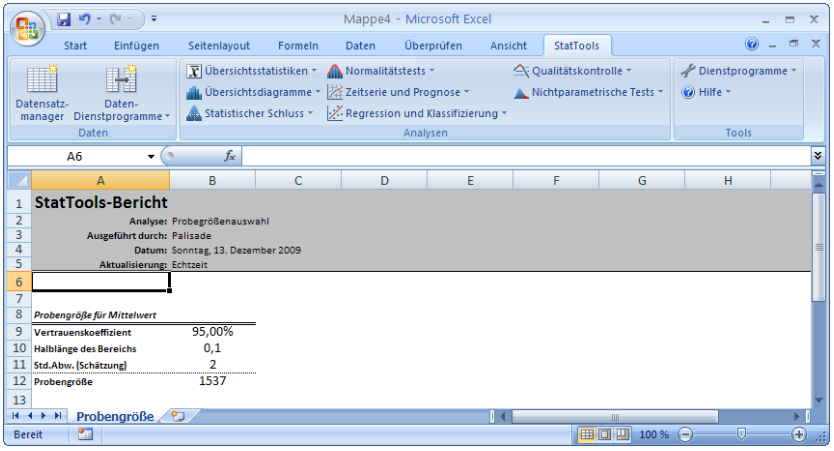


In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Zu schätzender Parameter** – wählt den Parametertyp aus, der aus der Werteprobe (deren Größe Sie derzeit bestimmen) beurteilt werden soll. Es sind die Parametertypen **Mittelwert, Proportion** (Werte zwischen 0 und 1), **Mittelwertdifferenzen** und **Proportionsdifferenzen** verfügbar.

- **Vertrauensintervallangabe** – hier hängen die Optionen (wie gezeigt) vom ausgewählten zu schätzenden Parameter ab:
  - Wenn es sich bei diesem Parameter um den **Mittelwert** und die **Mittelwertdifferenzen** handelt, müssen Sie zuerst den gewünschten **Vertrauenskoeffizienten** (gewöhnlich zwischen 90 % und 100 %), die **Halblänge des Bereichs** (d.h., die Plus- oder Minuskomponente des Bereichs) und die **geschätzte Standardabweichung** der Population eingeben. Hinweis: Vertrauenskoeffizient und Bereichslänge stehen im Zusammenhang, da bei einem höheren Vertrauenskoeffizienten ein längeres Vertrauensintervall erforderlich ist.
- Wenn es sich bei dem zu schätzenden Parameter um die **Proportion** und die **Proportionsdifferenzen** handelt, müssen Sie zuerst den gewünschten **Vertrauenskoeffizienten** (gewöhnlich zwischen 90 % und 100 %), die **Halblänge des Bereichs** (d.h., die Plus- oder Minuskomponente des Bereichs) und die **geschätzte Proportion** (d.h., einen Wert zwischen 0 und 1) eingeben. Falls die **Proportionsdifferenzen** geschätzt werden sollen, müssen Sie für jede Population eine **geschätzte Proportion** angeben.

Bericht über Probegrößen-schätzung



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – unwichtig
- **Mit Daten verknüpfen** – unwichtig

Befehl *Einseitige ANOVA*

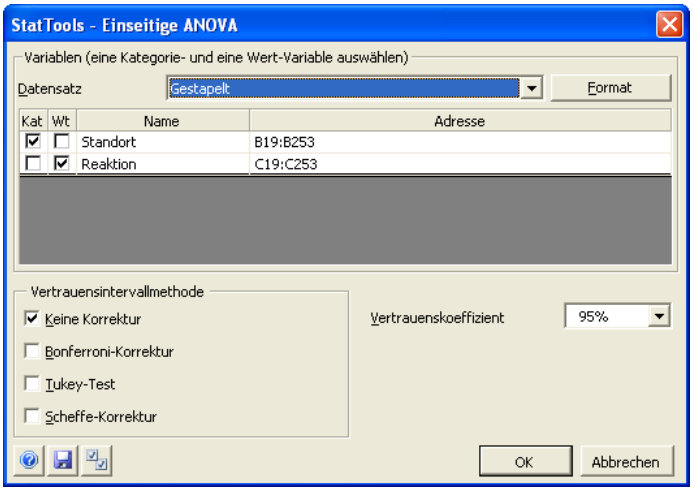
Führt eine einseitige ANOVA der Variablen aus

Bei diesem Befehl handelt es sich um eine Verallgemeinerung des Vorgangs mit zwei Werteproben beim Vergleich von Mittelwerten zwischen zwei Populationen. Bei der einseitigen ANOVA werden die Mittelwerte aus mindestens zwei (und oft mehr als zwei) Populationen verglichen. Das wird mithilfe einer ANOVA-Tabelle ausgeführt. Durch diese Tabelle werden eigentlich zwei Variationsquellen verglichen, d.h., die Variation innerhalb jeder Population wird mit der Variation unter den Probemittelwerten aus den verschiedenen Populationen verglichen. Ist die letztere Variation im Vergleich zur ersteren relativ hoch (was durch einen F-Test festgestellt werden kann), ist das ein Beweis dafür, dass zwischen den Populations-Mittelwerten erhebliche Differenzen bestehen.

Der wichtigste Wert in der ANOVA-Tabelle ist der *p*-Wert. Ein geringer *p*-Wert ist ein Beweis für unterschiedliche Populations-Mittelwerte. Außer der ANOVA-Tabelle sollten Sie sich aber auch die Vertrauensintervalle ansehen, um alle Differenzen zwischen den Mittelwertspaaren erkennen zu können. Vertrauensintervalle, die *nicht* den Wert 0 einschließen, sind ein Beweis dafür, dass die Mittelwerte *ungleich* sind. StatTools bietet mehrere Arten von Vertrauensintervallen, die jeweils auf einer etwas anderen Methode beruhen.

Dialogfeld ANOVA

Zum Definieren dieser Analyse wird das Dialogfeld **Einseitige ANOVA** verwendet.

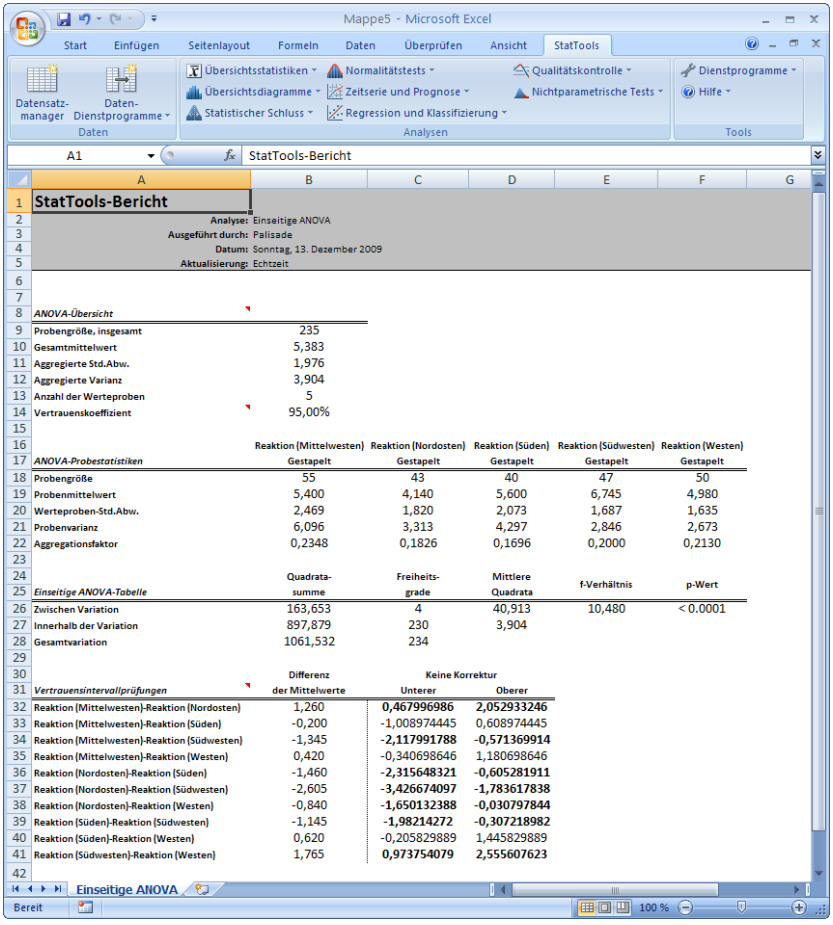


Für diese Analyse müssen zwei oder mehr Variablen ausgewählt werden. Der ausgewählten Datensätze können aus gestapelten oder entstapelten Daten bestehen. Diese Variablen können aus verschiedenen Datensätzen stammen.

Im Dialogfeld **Einseitige ANOVA** sind folgende Optionen verfügbar:

- **Vertrauensintervallmethode** – wählt eine oder mehrere der folgenden Methoden aus, um die Vertrauensintervalle für einzelne Variablen zu korrigieren: **Keine Korrektur**, **Bonferroni-Korrektur**, **Tukey-Test** und **Scheffe**.
- **Vertrauenskoeffizient** – dies ist ein gleichzeitiger Vertrauenskoeffizient für die Ergebnisse aus allen Variablen. Mit anderen Worten, Sie wollen sicher sein, dass *alle* Vertrauensintervalle auch die entsprechenden Populations-Mittelwertdifferenzen enthalten. Aus technischen Gründen wird der effektive Gesamtvertrauenskoeffizient normalerweise geringer sein als dieser unter **Keine Korrektur** für die Vertrauensintervalle angegebene Koeffizient. Das ist der Grund für die vorstehend genannten Korrekturmethoden. Dadurch werden die Längen der Vertrauensintervalle entsprechend korrigiert oder erweitert, sodass der Gesamtvertrauenskoeffizient dann dem angegebenen entspricht.

Bericht über einseitige ANOVA



StatTools-Bericht

Analyse: Einseitige ANOVA
Ausgeführt durch: Palisade
Datum: Sonntag, 13. Dezember 2009
Aktualisierung: Echtzeit

ANOVA-Übersicht

Probengröße, insgesamt	235
Gesamtmittelwert	5,383
Aggregierte Std.Abw.	1,976
Aggregierte Varianz	3,904
Anzahl der Werteproben	5
Vertrauenskoeffizient	95,00%

ANOVA-Probestatistiken	Reaktion (Mittelwesten) Gestapelt	Reaktion (Nordosten) Gestapelt	Reaktion (Süden) Gestapelt	Reaktion (Südwesten) Gestapelt	Reaktion (Westen) Gestapelt
Probengröße	55	43	40	47	50
Probenmittelwert	5,400	4,140	5,600	6,745	4,980
Werteproben-Std.Abw.	2,469	1,820	2,073	1,687	1,635
Probenvarianz	6,096	3,313	4,297	2,846	2,673
Aggregationsfaktor	0,2348	0,1826	0,1696	0,2000	0,2130

Einseitige ANOVA-Tabelle	Quadrata-summe	Freiheits-grade	Mittlere Quadrata	f-Verhältnis	p-Wert
Zwischen Variation	163,653	4	40,913	10,480	< 0.0001
Innerhalb der Variation	897,879	230	3,904		
Gesamtvariation	1061,532	234			

Vertrauensintervallprüfungen	Differenz der Mittelwerte	Keine Korrektur Unterer	Keine Korrektur Oberer
Reaktion (Mittelwesten)-Reaktion (Nordosten)	1,260	**0,467996986**	**2,052933246**
Reaktion (Mittelwesten)-Reaktion (Süden)	-0,200	-1,008974445	0,608974445
Reaktion (Mittelwesten)-Reaktion (Südwesten)	-1,345	**-2,117991788**	**-0,571369914**
Reaktion (Mittelwesten)-Reaktion (Westen)	0,420	-0,340698646	1,180698646
Reaktion (Nordosten)-Reaktion (Süden)	-1,460	**-2,315648321**	**-0,605281911**
Reaktion (Nordosten)-Reaktion (Südwesten)	-2,605	**-3,426674097**	**-1,783617838**
Reaktion (Nordosten)-Reaktion (Westen)	-0,840	**-1,650132388**	**-0,030797844**
Reaktion (Süden)-Reaktion (Südwesten)	-1,145	**-1,98214272**	**-0,307218982**
Reaktion (Süden)-Reaktion (Westen)	0,620	-0,205829889	1,445829889
Reaktion (Südwesten)-Reaktion (Westen)	1,765	**0,973754079**	**2,555607623**

In dem hier gezeigten einseitigen ANOVA-Bericht ist gleich oben die Übersichtsstatistik für jede Population (d.h., für jede Fertigungsanlage) angezeigt. Nach der Übersichtsstatistik ist eine Tabelle der Probenstatistiken für jede mit einbezogene Variable zu sehen. Danach folgt die ANOVA-Tabelle. Aus dem sehr geringen *p*-Wert in diesem Beispiel geht ohne Zweifel hervor, dass die Mittelwert-Punktwerte für die fünf Fertigungsanlagen nicht alle gleich sind. Aus den Vertrauensintervallen ganz unten im Bericht ist zu erkennen, welche Mittelwerte sich von welchen anderen Mittelwerten unterscheiden. Die Paare mit durch Fettdruck hervorgehobenen Werten haben erheblich unterschiedliche Mittelwerte.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

Befehl *Zweiseitige ANOVA*

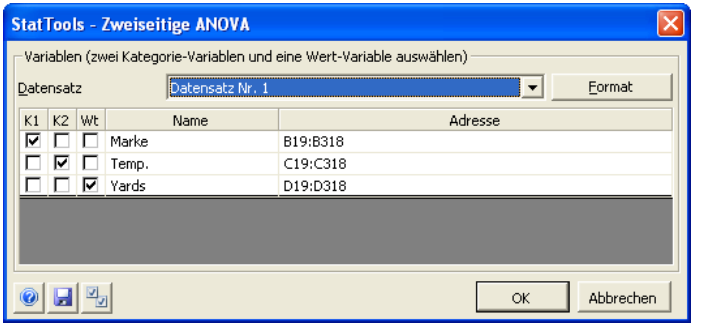
Führt eine zweiseitige ANOVA der Variablen aus

Durch diesen Befehl wird eine zweiseitige Varianzanalyse ausgeführt. diese Analyse wird gewöhnlich im Zusammenhang mit einer experimentellen Planung ausgeführt, die zwei Faktoren enthält, die beide auf mehrere „Behandlungsstufen“ eingestellt sind Bei der Golfballbewertung könnte es sich beispielsweise um die zwei Faktoren *Marke* und A*ußentemperatur* handeln. In diesem Fall würden für *Marke* dann die Behandlungsstufen *A* bis *E* und für *Temperatur* die Behandlungsstufen *Kühl, Mild* und *Warm* verwendet werden. Als Wertvariable würde *Balltreiblänge (Yard)* eingesetzt und die sich daraus ergebenden Beobachtungen würden dann für eine Anzahl von Golfbällen der einzelnen *Marken/Temperatur*-Kombinationen erfasst werden. Durch diese Untersuchung soll herausgefunden werden, ob erhebliche Mittelwertdifferenzen unter den verschiedenen Behandlungsstufenkombinationen bestehen.

Die Daten für die zweiseitige ANOVA *müssen* gestapelt sein. Mit anderen Worten, es müssen zwei Kategorievariablen (entsprechend der im vorstehenden Beispiel genannten Kategorien *Marke* und *Außentemperatur*) und auch eine Wertvariable (siehe *Balltreiblänge*) vorhanden sein. Außerdem müssen die Daten „ausgeglichen“ sein, d.h., jede Behandlungsstufenkombination sollte aus der gleichen Anzahl an Beobachtungen bestehen. Natürlich können auch unausgeglichene Daten analysiert werden, aber das sollte am besten durch Regression (d.h., mittels Pseudovariablen) geschehen.

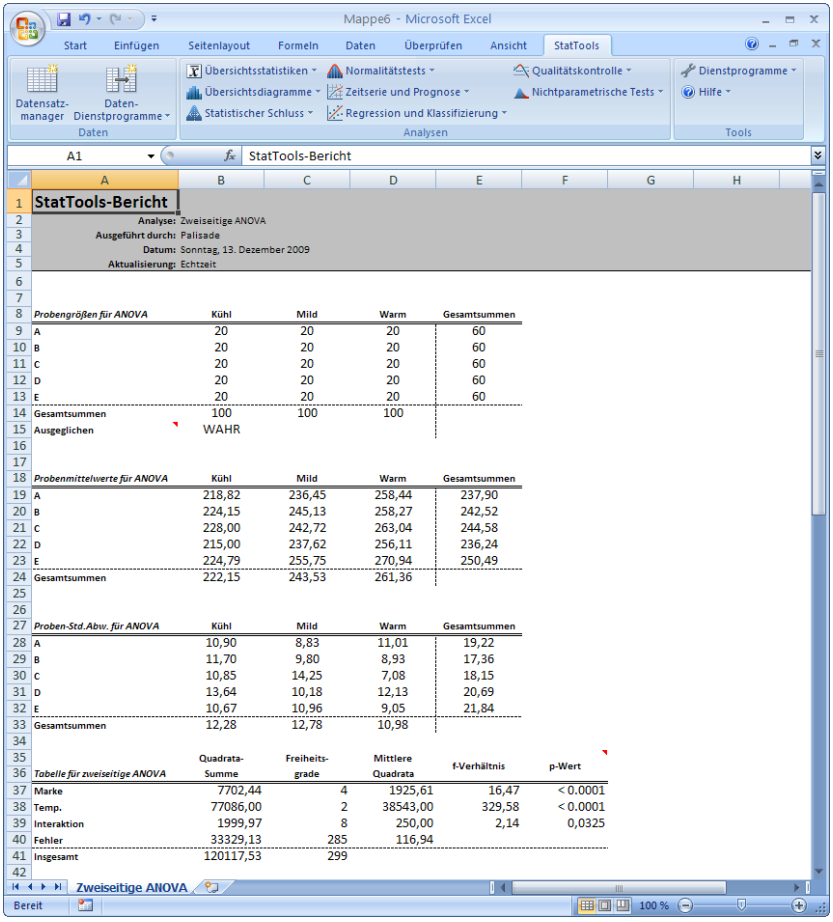
***Dialogfeld* ANOVA**

Zum Definieren dieser Analyse wird das Dialogfeld **Zweiseitige ANOVA** verwendet.



Für diese Analyse müssen zwei Kategorievariablen (**K1** und **K2)** und eine Wertvariable (**Wert**) ausgewählt werden. Auch muss der ausgewählte Datensatz aus gestapelten Daten bestehen.

Bericht über zweiseitige ANOVA



Bei den obersten drei Elementen (Probengrößen, Probenmittelwerten und Probenstandardabweichungen) handelt es sich um Übersichtsfaktoren für die verschiedenen Behandlungsstufenkombinationen. Der untere Teil der Ausgabe besteht aus der ANOVA-Tabelle. In dieser Tabelle befinden sich drei wichtige *p*-Werte: zwei für „Hauptwirkungen“ und einer für „Interaktionen“. Aus den Hauptwirkungen geht hervor, ob erhebliche Mittelwertdifferenzen zwischen den Faktoren bestehen. Die Hauptwirkung **Temperatur** zeigt z.B., ob die Werte in den Zellen **B17** bis **D17** erhebliche Unterschiede aufweisen. (Das ist hier in der Tat der Fall, wie in der ANOVA-Tabelle aus dem geringen *p*-Wert für *Temperatur* zu sehen ist.)

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Da ausgeglichene Daten erforderlich sind, kann es hier keine fehlenden Daten geben.
- **Mit Daten verknüpfen –** Alle ANOVA-Formeln sind mit den Daten verknüpft. Falls sich die Daten ändern, werden die Ergebnisse automatisch entsprechend aktualisiert. Die Ausgeglichenheit des Experiments kann durch Ausführung der Analyse bestätigt werden. Datenänderungen können die Ausgeglichenheit beeinträchtigen und ungültige Ergebnisse verursachen.

Befehl *Chi-Quadrat-Unabhängigkeitstest*

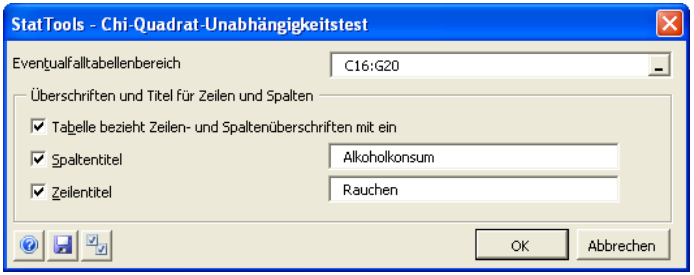
Überprüft die Unabhängigkeit unter den Attributen in den Zeilen und Spalten einer Eventualfalltabelle

Mithilfe dieses Befehl können Sie überprüfen, ob die Attribute in den Zeilen und Spalten einer Eventualtabelle unabhängig voneinander sind. Wenn in der Eventualfalltabelle beispielsweise Personengruppen in verschiedenen Alkoholkonsum- und Raucherkategorien aufgelistet sind, kann durch dieses Verfahren getestet werden, ob sich Raucherangewohnheiten unabhängig von Alkoholkonsumangewohnheiten entwickeln. Bei der Eventualfalltabelle (die auch Kreuztabelle genannt wird) kann es sich auch um die in Excel verwendete Pivot-Tabelle handeln,

die jedoch etwas anders funktioniert als die meisten Verfahren in StatTools. In StatTools ist für diese Analyse nur eine rechteckige Eventualfalltabelle erforderlich. In dieser Tabelle sollte jede Zelle einen Zählungswert für Beobachtungen in einer bestimmten Zeilen-/Spaltenkombination enthalten (z.B. für die Kombination **Abstinenzler/starke Raucher**). Die Tabelle kann mit Überschriften für Zeilen und Spalten versehen sein und auch die Gesamtsummen für Zeilen und Spalten aufweisen, aber diese Einzelheiten sind nicht unbedingt erforderlich und sorgen nur dafür, dass die StatTools-Berichte dann vielleicht etwas übersichtlicher sind.

***Dialogfeld* Chi-Quadrat-Unabhängigkeitstest**

Zum Definieren dieser Analyse wird das Dialogfeld **Chi-Quadrat-Unabhängigkeitstest** verwendet:

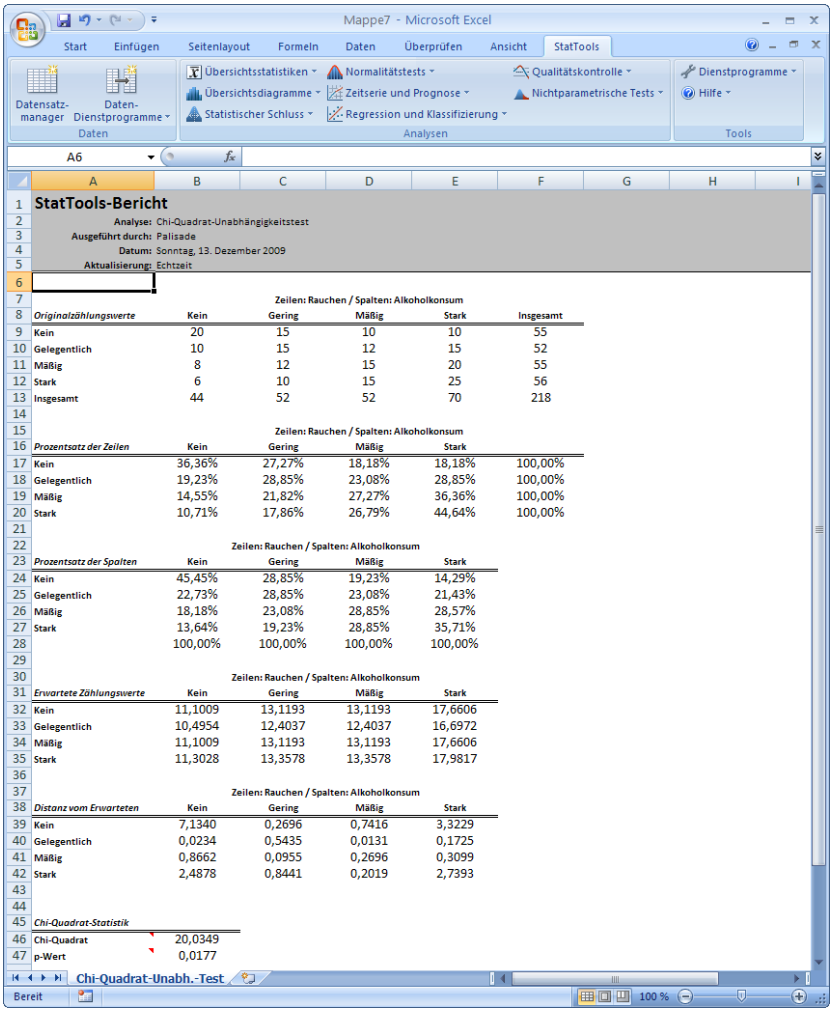


In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Überschriften und Titel für Zeilen und Spalten** – wählt eine oder mehrere von folgenden Optionen aus: **Tabelle bezieht Zeilen- und Spaltenüberschriften mit ein**, d.h. Überschriften in der ganz links befindlichen Spalte und der obersten Zeile der Tabelle; **Spaltentitel**, d.h. der Titel, der die Spalten in der

Tabelle beschreiben soll; **Zeilentitel**, d.h., der Titel, den Sie für die Zeilen in der Tabelle verwenden wollen.

Bericht über Chi-Quadrat-Unabhängigkeitstest



StatTools-Bericht

Analyse: Chi-Quadrat-Unabhängigkeitstest
Ausgeführt durch: Palisade
Datum: Sonntag, 13. Dezember 2009
Aktualisierung: Echtzeit

Zeilen: Rauchen / Spalten: Alkoholkonsum

Originalzählungswerte	Kein	Gering	Mäßig	Stark	Insgesamt
Kein	20	15	10	10	55
Gelegentlich	10	15	12	15	52
Mäßig	8	12	15	20	55
Stark	6	10	15	25	56
Insgesamt	44	52	52	70	218

Zeilen: Rauchen / Spalten: Alkoholkonsum

Prozentsatz der Zeilen	Kein	Gering	Mäßig	Stark	
Kein	36,36%	27,27%	18,18%	18,18%	100,00%
Gelegentlich	19,23%	28,85%	23,08%	28,85%	100,00%
Mäßig	14,55%	21,82%	27,27%	36,36%	100,00%
Stark	10,71%	17,86%	26,79%	44,64%	100,00%

Zeilen: Rauchen / Spalten: Alkoholkonsum

Prozentsatz der Spalten	Kein	Gering	Mäßig	Stark
Kein	45,45%	28,85%	19,23%	14,29%
Gelegentlich	22,73%	28,85%	23,08%	21,43%
Mäßig	18,18%	23,08%	28,85%	28,57%
Stark	13,64%	19,23%	28,85%	35,71%
	100,00%	100,00%	100,00%	100,00%

Zeilen: Rauchen / Spalten: Alkoholkonsum

Erwartete Zählungswerte	Kein	Gering	Mäßig	Stark
Kein	11,1009	13,1193	13,1193	17,6606
Gelegentlich	10,4954	12,4037	12,4037	16,6972
Mäßig	11,1009	13,1193	13,1193	17,6606
Stark	11,3028	13,3578	13,3578	17,9817

Zeilen: Rauchen / Spalten: Alkoholkonsum

Distanz vom Erwarteten	Kein	Gering	Mäßig	Stark
Kein	7,1340	0,2696	0,7416	3,3229
Gelegentlich	0,0234	0,5435	0,0131	0,1725
Mäßig	0,8662	0,0955	0,2696	0,3099
Stark	2,4878	0,8441	0,2019	2,7393

Chi-Quadrat-Statistik	
Chi-Quadrat	20,0349
p-Wert	0,0177

Der vorstehende Bericht zeigt das grundlegende Ergebnis des Tests, d.h., den *p*-Wert. Wenn dieser *p*-Wert (wie hier) sehr gering ist, können wir daraus schließen, dass die Attribute in den Zeilen und Spalten nicht unabhängig voneinander sind. Wir sollten uns die Zahlen auf diesem Blatt genauer ansehen, um besser zu verstehen, wie Raucher- und Alkoholkonsumangewohnheiten miteinander in Beziehung stehen.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – In den Zellen der Eventualfalltabelle sollte es keine fehlenden Daten geben.
- **Mit Daten verknüpfen** – Im Bericht über den Chi-Quadrat-Unabhängigkeitstest sind die Formeln mit den Daten verknüpft. Falls sich die Zählungswerte in der ursprünglichen Eventualfalltabelle ändern, werden die Ausgaben auf diesem Blatt automatisch entsprechend aktualisiert.

Menü *Normalitätstests*

Da in so vielen statistischen Vorgängen davon ausgegangen wird, dass der betreffende Datensatz normal verteilt ist, sollten Methoden vorhanden sein, um diese Annahme zu überprüfen. StatTools bietet drei diesbezügliche gebräuchliche Überprüfungen, die in diesem Abschnitt beschrieben werden.

Befehl *Chi-Quadrat-Normungstest*

Überprüft, ob die beobachteten Daten für eine Variable normal verteilt sind

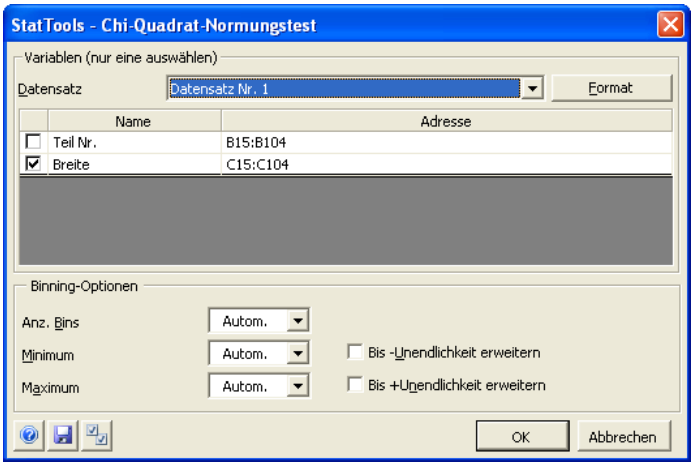
Bei diesem Testverfahren wird durch einen Chi-Quadrat-Anpassungsgütetest überprüft, ob die in einer bestimmten Variable beobachteten Daten aus einer Normalverteilung kommen konnten. Zu diesem Zweck wird unter Verwendung der von Ihnen angegebenen Kategorien ein Histogramm dieser Variable erstellt und dann durch ein Histogramm einer Normalverteilung überlagert. Wenn dann beide Histogramme praktisch gleichförmig sind, kann die Nullhypothese einer Normalanpassung nicht zurückgewiesen werden.

Der formale Test wird dadurch ausgeführt, dass die beobachteten Zählungswerte in den verschiedenen Kategorien mit den erwarteten Zählungswerten verglichen werden, die auf der Normalitätshypothese basieren. Dieser Vorgang ermöglicht Ihnen sogar, mehrere Variablen einzeln auf Normalität zu überprüfen. Es wird dann für jede von Ihnen ausgewählte Variable ein Histogramm erstellt und der Chi-Quadrat-Test ausgeführt.

Für diesen Test ist lediglich erforderlich, dass mindestens eine numerische Variable verfügbar ist. Des Weiteren empfehlen die meisten Analytiker, mindestens 100 Beobachtungen auszuführen – je mehr, um so besser!

Dialogfeld Chi-Quadrat-Normungstest

Zum Definieren dieser Analyse wird das Dialogfeld **Chi-Quadrat-Normungstest** verwendet:



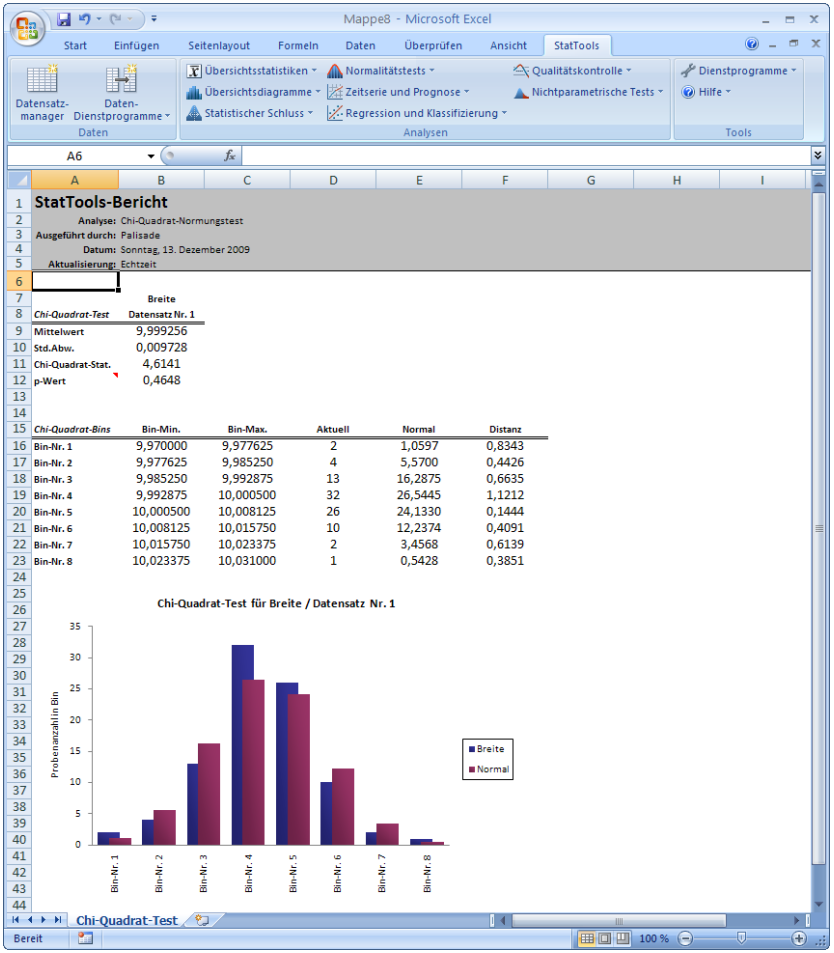
Für diesen Test kann eine Variable ausgewählt werden. Der ausgewählte Datensatz kann aus gestapelten oder entstapelten Daten bestehen.

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Anzahl Bins –** gibt die Anzahl der Bins an oder legt fest, dass die Anzahl automatisch berechnet werden soll.
- **Minimum** und **Maximum –** durch **autom.** wird angegeben, dass der Minimal- und Maximalwert des Datensatzes dazu benutzt werden soll, die Minimal- und Maximallänge der intervallgleichen Bins zu berechnen. Es kann jedoch das erste und letzte Bin hinzugefügt werden, indem die Optionen **Unendlich ins Negative fortlaufen lassen** und **Unendlich ins Positive fortlaufen lassen** verwendet werden. Sofern **autom.** nicht ausgewählt ist, kann ein bestimmter Minimal- oder Maximalwert eingegeben werden, um zu zeigen, wo die Bins beginnen und enden sollen. Dadurch haben Sie die Möglichkeit, einen bestimmten Bereich für das Binning einzugeben, ohne dabei die Minimal- und Maximalwerte des Datensatzes berücksichtigen zu müssen.
- **Bis -Unendlichkeit erweitern –** gibt an, dass das erste benutzte Bin vom angegebenen Minimalwert aus zur Minusseite hin ins Unendliche reichen soll. Alle anderen Bins werden in diesem Fall von gleicher Länge sein. Unter gewissen Umständen kann dadurch das Testen von Datensätzen mit unbekannter unterer Begrenzung verbessert werden.

- **Bis +Unendlichkeit erweitern –** gibt an, dass das letzte benutzte Bin vom angegebenen Maximalwert aus zur Plusseite hin ins Unendliche reichen soll. Alle anderen Bins werden in diesem Fall von gleicher Länge sein. Unter gewissen Umständen kann dadurch das Testen von Datensätze mit unbekannter oberer Begrenzung verbessert werden.

Bericht über Chi-Quadrat-Normungstest



Die Ergebnisse dieses Tests sind im vorstehenden Bericht zu sehen. Der *p*-Wert von 0,4776 ist ein guter Beweis dafür, dass die Summen normal verteilt sind. Ein weiterer Beweis hierfür ist in den Histogrammen in Abbildung 4 und in den Häufigkeitsdaten zu sehen. Dabei sollten jedoch zwei Dinge besonders beachtet werden: Wenn nicht genügend Beobachtungen (d.h., erheblich weniger als 100) vorhanden sind, kann durch den Chi-Quadrat-Test nicht gut zwischen Normalität und Anormalität unterschieden werden. Das bewirkt dann, dass der *p*-Wert gewöhnlich nicht gering genug ist, um die Normalitätshypothese zurückzuweisen. Im Wesentlichen sieht *fast alles* scheinbar normal aus, wenn die Datensätze zu klein sind. Wenn der Datensatz andererseits sehr groß ist (z.B. aus Hunderten von Beobachtungen besteht), ist der *p*-Wert gewöhnlich immer gering und weist damit scheinbar auf Anormalität hin. Das hat damit zu tun, dass bei großen Datensätzen jede kleine Unebenheit in der Kurve wahrscheinlich bereits einen kleinen *p*-Wert erstellt. In diesem Fall sollte ein *praktischer* Test (d.h., ein Erfahrungstest) ausgeführt werden, indem Sie sich fragen: „Sind die Histogramme, praktisch gesehen, wirklich so unterschiedlich?“

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Das Histogramm und alle Formeln für den Test sind mit den Originaldaten verknüpft. Wenn sich daher die Daten ändern, werden Histogramm und Testergebnisse automatisch entsprechend aktualisiert.

Befehl *Lilliefors-Test*

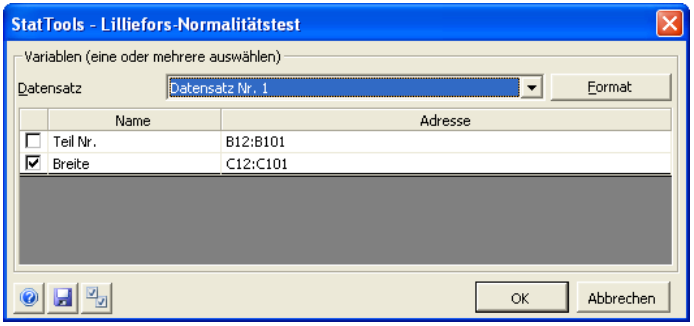
Überprüft, ob die beobachteten Daten für eine Variable normal verteilt sind

Dieser Test liefert einen leistungsstärkeren Normalitätstest als durch den bekannteren Chi-Quadrat-Anpassungsgütetest möglich ist („leistungsstärker“ bedeutet in diesem Fall, dass durch den Test besser irgendeine anormale Anpassung erkannt werden kann). Der Lilliefors-Test basiert auf einem Vergleich der erfahrungsgemäßen CDF mit einer normalen CDF (Summenverteilungsfunktion). Dadurch kann die Wahrscheinlichkeit erkannt werden, dass diese Summenverteilung nicht größer als ein bestimmter Wert ist.

Die erfahrungsgemäße CDF beruht auf den vorhandenen Daten. Wenn es sich beispielsweise um 100 Beobachtungen handelt und die 13. kleiner als 137 ist, hat die erfahrungsmäßige CDF (bei Auswertung auf Basis von 137) den Wert 0,13. Durch den Lilliefors-Test wird die maximale vertikale Distanz zwischen erfahrungsgemäßer und normaler CDF festgestellt. Dieses Maximum wird dann mit den tabellarisierten Werten verglichen, die auf der Probengröße basieren. Wenn die beobachtete maximale vertikale Distanz ausreichend groß ist, besteht der Beweis dafür, dass die Daten *nicht* aus einer Normalverteilung stammen.

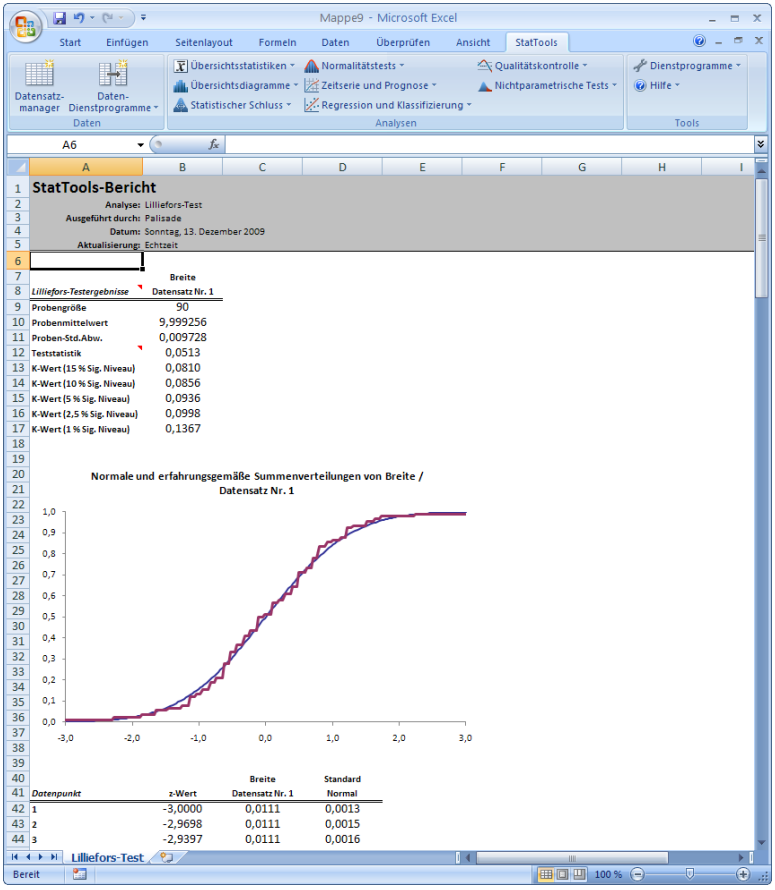
***Dialogfeld* Lilliefors-Normalitätstest**

Zum Definieren dieser Analyse wird das Dialogfeld **Lilliefors-Normalitätstest** verwendet:



Für diesen Test können eine oder mehrere Variablen ausgewählt werden. Ferner muss der ausgewählte Datensatz aus entstapelten Daten bestehen. Die Variablen können aber aus verschiedenen Datensätzen stammen.

Lilliefors-Testbericht



Die Ergebnisse dieses Tests sind im vorstehenden Bericht zu sehen. Genau wie in den meisten Hypothesetests ist auch hier kein *p*-Wert vorhanden, aber wir können aus der Anweisung erkennen, dass die maximale vertikale Distanz groß genug ist, um die Normalitätshypothese anzuzweifeln. Ein weiterer Beweis hierfür geht aus den CDFs im beigefügten Diagramm hervor. Die beiden Kurven scheinen eigentlich einander ziemlich gut angepasst zu sein, sodass es vielleicht in der Praxis ausreicht. Mit anderen Worten, wir könnten daraus schließen, dass diese Daten für unsere Zwecke so gut wie normal verteilt sind.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Die CDFs und alle Formeln für den Test sind mit den Originaldaten verknüpft. Wenn sich daher die Daten ändern, werden Diagramm und Testergebnisse automatisch entsprechend aktualisiert.

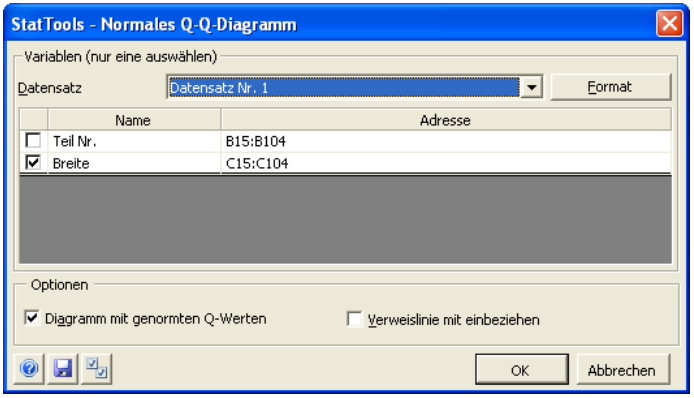
Befehl *Normales Q-Q-Diagramm*

Überprüft, ob die beobachteten Daten für eine Variable normal verteilt sind

Mithilfe dieses Befehls können Sie ein Quantil-Quantil-Diagramm für eine einzelne Variable erstellen. Daraus ergibt sich ein informeller Normalitätstest Obwohl die Einzelheiten vielleicht etwas kompliziert sind, ist das Ziel einfach, Die Quantile (oder Perzentile) aus den Daten mit den Quantilen aus einer Normalverteilung zu vergleichen. Falls die Daten im Wesentlichen normal sind, sollten die Punkte im Q-Q-Diagramm ungefähr eine 45-Grad-Linie ergeben. Offensichtliche Krümmung im Diagram mist jedoch ein Anzeichen von irgendeiner Anormalität (z.B. von einer gewissen Schiefe).

***Dialogfeld* Normales Q-Q-Diagramm**

Zum Definieren dieser Analyse wird das Dialogfeld **Normales Q-Q-Diagramm** verwendet:

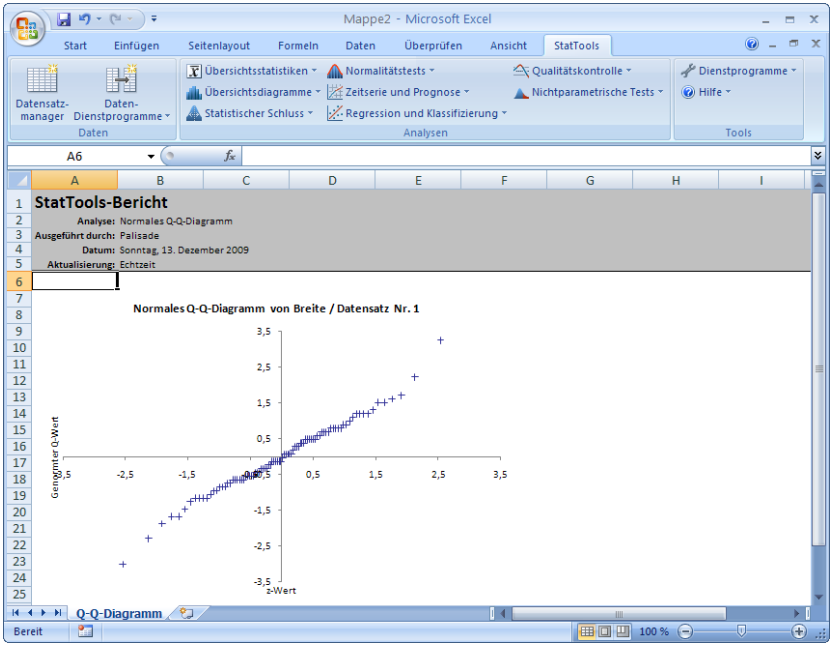


Für dieses Diagramm kann eine Variable ausgewählt werden. Der ausgewählte Datensatz muss aber aus entstapelten Daten bestehen.

In diesem Dialogfeld ist folgende Option zu finden:

- **Diagramm mit genormten Q-Werten** – legt fest, dass auf der y-Achse des Diagramms ein genormter Q-Wert anstelle von Q-Q-Daten verwendet werden soll. Dadurch sind Vergleiche der y-Achsenwerte der verschiedenen normalen Q-Q-Diagramme möglich.

Bericht über Normales Q-Q-Diagramm



Wie bereits erwähnt, ist dies ein *informeller* Normalitätstest. Es ist schwierig, genau zu bestimmen, wie genau in diesem Diagramm eine 45-Grad-Linie erreicht werden muss, um die Normalitätshypothese zu akzeptieren. Typischerweise suchen wir nach einer offensichtlichen Krümmung im Diagramm, die aber hier nicht zu finden ist.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Die Diagramme und alle Formeln sind mit den Daten verknüpft. Falls sich die Daten ändern, wird das Diagramm automatisch entsprechend aktualisiert.

Menü *Zeitserie und Prognose*

Die Befehle in diesem Menü haben mit der Analyse von Daten zu tun, die über längere Zeitperioden hinweg erfasst wurden, und zwar mittels Anwendungen für Prognosen und Qualitätskontrolle. Die verfügbaren Prognosemethoden schließen gleitende Durchschnittsmethode, einfaches exponentielles Glätten, exponentielle Glättungsmethode für erfasste Tendenz (Holt) und exponentielle Glättungsmethode für erfasste Saisonabhängigkeit (Winters) mit ein.

Befehl *Zeitserien-Diagramm*

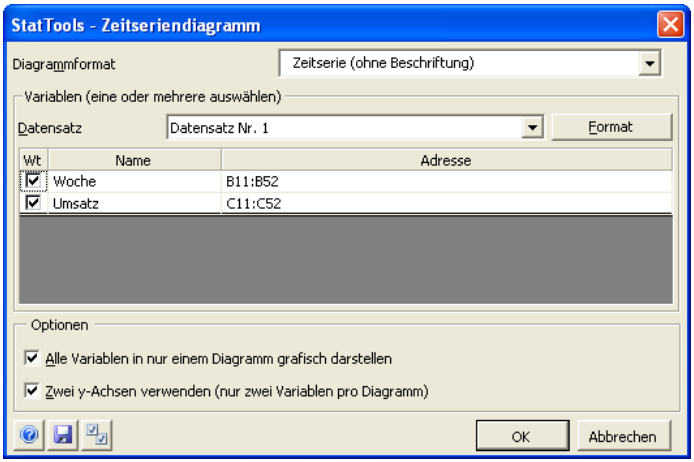
Erstellt ein Zeitserien-Diagramm für Variablen

Über diesen Befehl können eine oder auch mehrere Zeitserien-Variablen in ein und demselben Diagramm grafisch dargestellt werden. Falls zwei Variablen ausgewählt werden, können Sie auf der y-Achse dieselbe oder auch eine unterschiedliche Skalierung für die beiden Variablen verwenden. Die unterschiedliche Skalierung ist zu empfehlen, wenn die Wertbereiche für die beiden Variablen sich erheblich unterscheiden. Wenn jedoch mehr als zwei Variablen grafisch dargestellt werden, müssen sie alle die gleiche vertikale Skala verwenden.

Auch muss sich mindestens eine numerische Variable im Datensatz befinden. Ferner kann eine Datums-Variable vorhanden sein, aber wenn diese zum Bezeichnen der horizontalen Achse des Diagramms verwendet werden soll, muss sie als Bezeichnungs- oder „Label"-Variable ausgewählt werden.

Dialogfeld Zeitserien-Diagramm

Zum Definieren dieses Diagrammtyps wird das Dialogfeld **Zeitserien-Diagramm** verwendet:

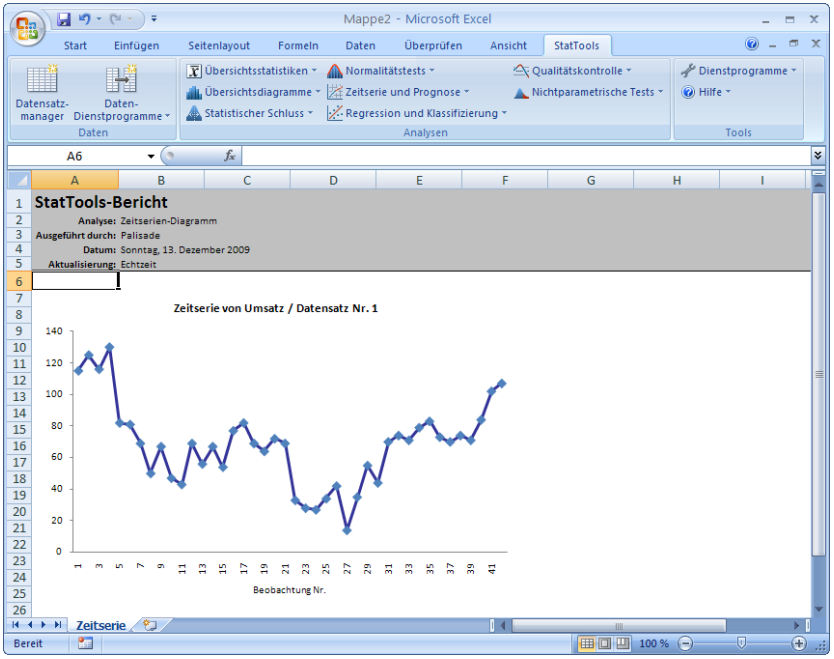


Für die grafische Darstellung können eine oder mehrere Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können aber aus verschiedenen Datensätzen stammen. Die Label-Variable (Kontrollkästchen **LBL**) ist auf der x-Achse zu sehen.

Unter **Optionen** ist in diesem Dialogfeld Folgendes zu finden:

- **Alle Variablen in nur einem Diagramm grafisch darstellen –** wählt alle Variablen für ein und dasselbe Diagramm aus.
- **Zwei y-Achsen verwenden –** wählt eine separate *y*-Achse für jede Variable zur Anzeige in einem Zwei-Variablen-Diagramm aus. Für jede Variable können dann Einheiten und Werte im Diagramm angezeigt werden.

Zeitserien-Diagramm (zwei Variablen)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Diagramm ist mit den Daten verknüpft. Falls sich die Daten ändern, wird das Diagramm daher automatisch entsprechend aktualisiert.

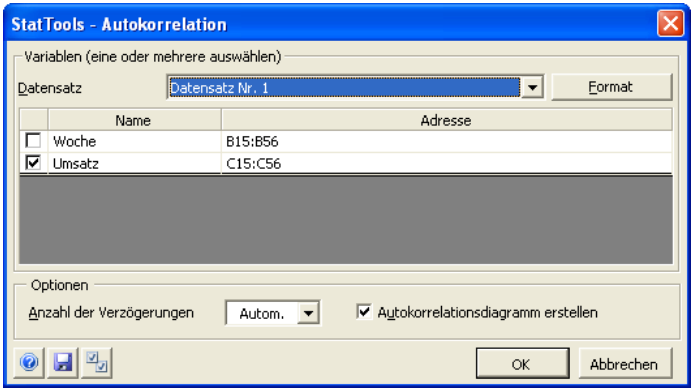
Befehl *Autokorrelation*

Berechnet die Autokorrelationen für Variablen

Mithilfe dieses Befehls kann die Autokorrelation für die ausgewählte numerische Variable berechnet werden. Gewöhnlich ist dies eine Zeitserien-Variable, aber StatTools kann die Berechnungen auch für jede beliebige andere Variable ausführen. Sie können die gewünschte Anzahl an Autokorrelationen (d.h., die Anzahl der Verzögerungen) auswählen. Auch können Sie ein Diagramm der Autokorrelationen anzeigen lassen, das „Korrelogramm" genannt wird. Aus der Ausgabe geht dann hervor, ob irgendeine Autokorrelation erheblich vom Nullwert abweicht.

***Dialogfeld* Autokorrelation**

Zum Definieren dieses Diagrammtyps wird das Dialogfeld **Autokorrelation** verwendet:



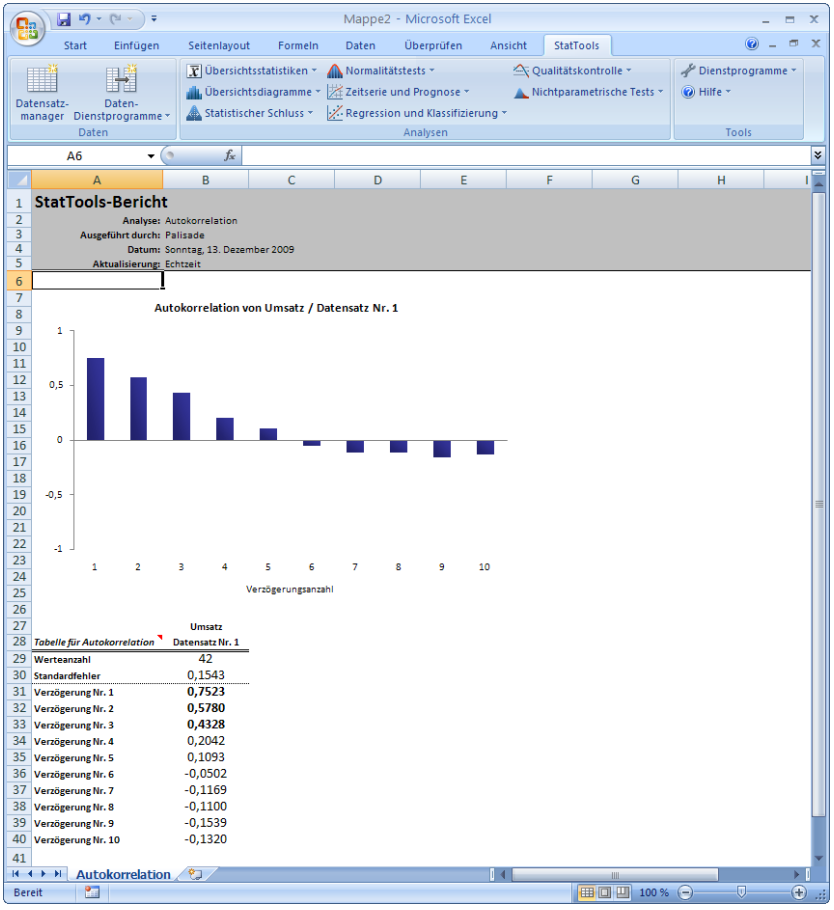
Für diese Analyse können mehrere Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln.

Unter **Optionen** ist in diesem Dialogfeld Folgendes zu finden:

- **Anzahl der Verzögerungen** – Dies ist die Anzahl der Verzögerungsperioden beim Berechnen der Autokorrelationen. Wenn **Autom**. ausgewählt ist, legt StatTools die passende Anzahl an Verzögerungen fest. Falls eine bestimmte Anzahl von Verzögerungen eingegeben wird, darf diese Anzahl nicht höher als 25 % der gesamten Beobachtungen in der Serie liegen. Wenn Sie beispielsweise 80 monatliche Werte haben, können Sie bis zu 20 Verzögerungen eingeben.

- **Autokorrelationsdiagramm erstellen –** erstellt ein Balkendiagramm, in dem die Höhe der einzelnen Balken der entsprechenden Autokorrelation entspricht.

Bericht über Autokorrelation



Dieser Bericht ist vorstehend zu sehen. Für jede Verzögerung wird die entsprechende Autokorrelation gezeigt, und zwar zusammen mit dem *ungefähren* Standardfehler.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Bei diesem Vorgang können fehlende Daten zu Anfang der Zeitserie gehandhabt werden, aber nicht mehr später in der Serie.

- **Mit Daten verknüpfen** – In StatTools ist die Ausgabe mit den Daten verbunden. Falls sich die Daten ändern, werden daher die Autokorrelationen (und das Korrelogramm) automatisch aktualisiert.

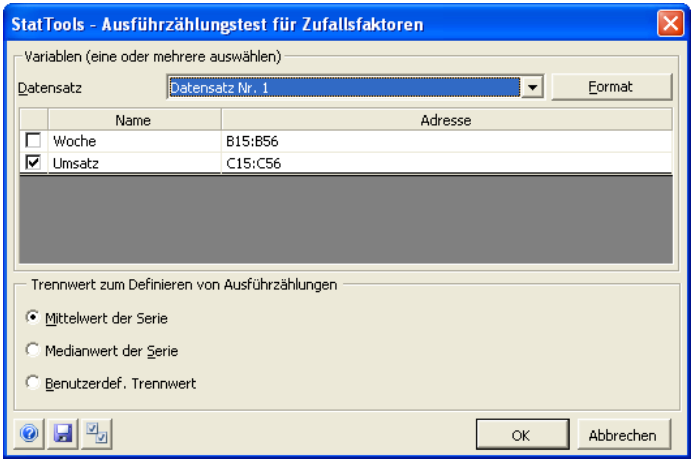
Befehl *Ausführzählungstest für Zufallsfaktoren*

Führt einen Ausführzählungstest aus, um festzustellen, ob es sich um eine Zufallsvariable handelt

Dieser Befehl ermöglicht Ihnen, in einer Variable die Zufallsfaktoren einer Werteserie zu überprüfen. Meistens handelt es sich dabei um eine Zeitserien-Variable. Auf diese Weise kann erkannt werden, wie viele Ausführ- oder Wertzählungen in der Serie enthalten sind. Eine „Ausführzählung" stellt dabei eine fortlaufende Anzahl von Werten auf der einen oder anderen Seite eines Trennpunktes (wie z.B. eines Mittelwerts oder Medianwerts der Serie) dar. Bei einer Zufallsserie würden wir weder mit zu wenigen noch mit zu vielen Ausführzählungen rechnen. Durch den Ausführzählungstest wird die Anzahl der Ausführungen gezählt und dann ein *p*-Wert für den Test angezeigt. Falls dieser *p*-Wert gering ist, können wir daraus schließen, dass es sich wahrscheinlich um keine Zufallsserie handelt. Das heißt, es sind weder zu viele noch zu wenige Ausführzählungen vorhanden.

***Dialogfeld* Ausführ-zählungstest für Zufallsfaktoren**

Zum Definieren dieser Analyse wird das Dialogfeld **Ausführzählungstest für Zufallsfaktoren** verwendet:

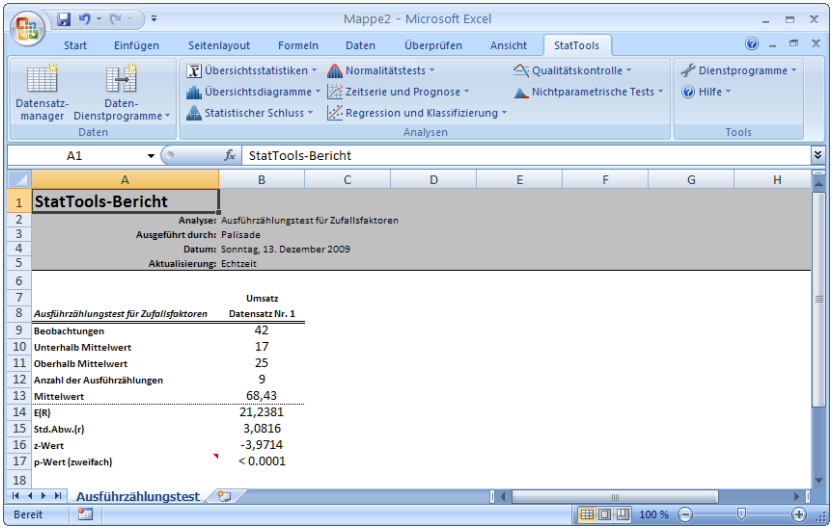


Für diese Analyse können mehrere Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können aber aus verschiedenen Datensätzen stammen.

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Trennwert zum Definieren von Ausführzählungen** – Ein Ausführzählungstest basiert immer auf der Zählung oberhalb oder unterhalb eines Trennwerts. Es kann sich dabei um den **Mittelwert der Serie** bzw. den **Medianwert der Serie** oder auch um irgendeinen anderen benutzerdefinierten Trennwert handeln, den Sie gern eingeben möchten.

Bericht über Ausführzählungs-test für Zufallsfaktoren



Der vorstehende Bericht zeigt die Anzahl der Ausführzählungen und die erwartete Anzahl an zufälligen Zählungen [E(R)]. Da 20 erheblich weniger als 24,8333 ist, kann man hier nicht von einer vollkommen zufälligen Serie sprechen. Das heißt, der Umsatz folgt nicht der Zickzacklinie, die durch eine Zufallsserie verursacht werden würde.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Bei diesem Vorgang können fehlende Daten zu Anfang und Ende der Zeitserie gehandhabt werden, aber nicht in der Mitte der Serie.

- **Mit Daten verknüpfen** – In StatTools ist die Ausgabe mit den Daten verbunden. Falls sich die Daten ändern, werden die Berichte daher automatisch entsprechend aktualisiert.

Befehl *Prognose*

Erstellt Prognosen für Zeitserien-Variablen

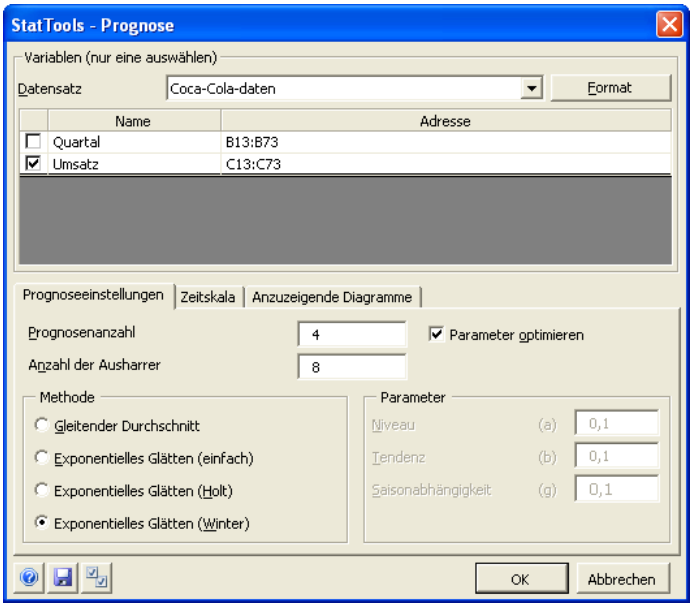
Wenn Sie diesen Befehl verwenden, stehen Ihnen mehrere Methoden zur Verfügung, um Prognosen für eine Zeitserien-Variable zu stellen. Diese Methoden schließen gleitende Durchschnittsmethode, einfaches exponentielles Glätten, exponentielle Glättungsmethode für erfasste Tendenz (Holt) und exponentielle Glättungsmethode für erfasste Saisonabhängigkeit (Winters) mit ein. Der Befehl **Prognose** ermöglicht Ihnen auch, die Daten vorweg saisonunabhängig zu machen, indem die Methode **Verhältnis zu gleitenden Durchschnitten** und ein vielfaches Saisonalitätsmodel verwendet werden. Anschließend können Sie dann irgendeine der Prognosemethoden (außer der Winter-Methode) verwenden, um Prognosen für die saisonunabhängigen Daten zu stellen und dann die Prognosen wieder saisonabhängig zu machen, damit Sie zurück zu den Originaleinheiten gelangen.

In den Prognoseberichten sind Spalten enthalten, in denen Sie die verschiedenen Berechnungen (z.B. für Glättung bei erfasster Tendenz (Holt), für saisonbedingte Faktoren bei Verhältnis zu gleitenden Durchschnitten usw.) sowie die Prognosen und auch die Prognosenfehler sehen können. Übersichtsfaktoren (MAE, RMSE und MAPE) sind ebenfalls mit einbezogen, um die Anpassung des Modells auf die beobachteten Daten verfolgen zu können. (Bei Verwendung einer der exponentiellen Glättungsmethoden können Sie über die Optimierung nach den Glättungskonstanten suchen, durch die RMSE minimiert wird.)

Außerdem sind mehrere Zeitserien-Diagramme verfügbar, einschließlich eines Diagramms der Originalserie, der Serie mit überlagerten Prognosen und der Prognosenfehler. Wenn die Daten saisonunabhängig gemacht werden sollen, sind diese Diagramme sowohl für die Originalserie als auch für die saisonunabhängige Serie verfügbar.

Dialogfeld Prognose

Zum Definieren von Prognosen wird das Dialogfeld **Prognose** verwendet:



Für diese Analyse kann eine Variable ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln.

***Prognoseoptionen – Dialogfeld* Prognose**

Über die Optionen in diesem Dialogfeld werden die zu verwendende Prognosenmethode und die dafür notwendigen Einstellungen angegeben. Es handelt sich dabei um folgende Optionen:

- **Prognosenanzahl** – legt die Anzahl der *zukünftigen* Perioden fest, für die Prognosen erstellt werden sollen.
- **Anzahl der Ausharrer** – legt die Anzahl der Beobachtungen fest, die in diesem Prognosemodell keine Verwendung finden sollen. Es können entweder alle Beobachtungen für das Prognosemodell verwendet werden (0 Ausharrer) oder es können einige zur Validierung oder Überprüfung zurückgehalten werden. Im letzteren Fall wird im Modell mithilfe der „nicht ausharrenden“ Beobachtungen prognostiziert und werden auf diese Weise die ausharrenden Beobachtungen vorausgesagt.

- **Parameter optimieren (nur exponentielle Glättungsmethoden) –** sucht nach der Glättungskonstanten, durch die RMSE für die nicht ausharrende Periode minimiert wird. Für die Optimierung müssen die im Dialogfeld angegebenen Parameter zwischen 0 und 1 liegen. Falls Sie die Parameterwerte direkt im Prognosebericht bearbeiten, müssen Sie darauf achten, dass nur Werte in diesem Bereich eingegeben werden.
- **Saisonunabhängig machen –** macht die Daten vor der Prognose saisonunabhängig. Bei saisonabhängigen Daten, d.h., bei Daten, die einem saisonbedingten Muster folgen, haben Sie zwei Möglichkeiten: Sie können die **Winter**-Methode verwenden, die sich direkt mit der Saisonabhängigkeit befasst, oder Sie können diese Option dazu benutzten, die Daten vorweg saisonunabhängig zu machen, und zwar mithilfe der Methode **Verhältnis zu gleitenden Durchschnitten**. Anschließend kann dann *jede beliebige* Methode dazu verwendet werden, eine Prognose für die saisonunabhängig gemachte Serie zu stellen.
- **Methode –** wählt die zu verwendende Prognosemethode aus. Es kann sich dabei um **Gleitenden Durchschnitt** oder **Einfach** bzw. das exponentielle Glätten von **Holt** oder **Winter** handeln.
- **Parameter –** legt die Parameter fest, die für die ausgewählte Prognosemethode verwendet werden sollen:
  - **Spanne** (nur bei gleitendem Durchschnitt). Es handelt sich hier um die Anzahl der aufeinander folgenden Beobachtungen, die für jeden gleitenden Durchschnitt verwendet werden.
  - **Niveau** (bei allen exponentiellen Glättungsmethoden). Dies ist ein Glättungsparameter, der im Wert zwischen 0 und 1 liegen kann. Der Standardwert für diesen Parameter ist 0,1.
  - **Tendenz** (bei der exponentiellen Glättungsmethode von **Holt** und **Winter)**. Dies ist ein zweiter Glättungsparameter, der ebenfalls im Wert zwischen 0 und 1 liegen kann. Der Standardwert für diesen Parameter ist 0,1.
  - **Saisonabhängigkeit** (nur bei der exponentiellen Glättungsmethode von Winter). Dies ist ein dritter Glättungsparameter, der wiederum im Wert zwischen 0 und 1 liegen kann. Der Standardwert für diesen Parameter ist 0,1.

Hinweis: Falls* Parameter optimieren *ausgewählt ist, können die Parameter für* Niveau, Tendenz *und* Saisonabhängigkeit *nicht eingestellt werden, da dies die Parameter sind, deren Werte optimiert werden sollen.

Zeitskala-Optionen – Dialogfeld Prognose

Über die Zeitskala-Optionen können die Zeiteinteilungs- und Zeitskala-Bezeichnungen für die analysierte Variable angegeben werden. Es handelt sich dabei um folgende Optionen:

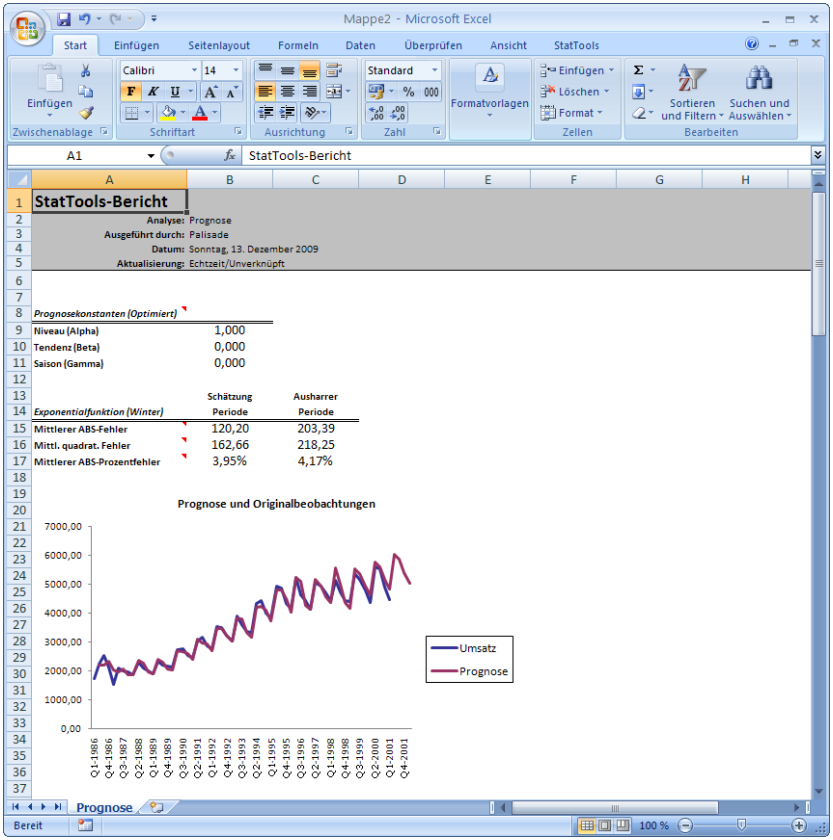
- **Saisonperiode –** legt den Typ der Zeitseriendaten fest. Es kann sich dabei um **Jährlich, Vierteljährlich, Monatlich, Wöchentlich, Täglich** oder **Kein(e)** handeln. Dies wird dazu verwendet, die Daten saisonabhängig zu machen und zu bezeichnen.
- **Beschriftungsart –** legt fest, wie die Zeitskala in den erstellten Diagrammen beschriftet werden soll.
- **Anfangsbeschriftung** – Legt den Eintrag für die erste Zeitskalenbeschriftung in dem Diagramm fest.

Diagrammoptionen – Dialogfeld Prognose

Über die Diagrammoptionen kann festgelegt werden, welche Prognosediagramme erstellt werden sollen. Folgende Diagramme sind verfügbar:

1) **Prognoseüberlagerung –** Dies ist ein Zeitseriendiagramm der durch die Prognose erstellten Datenwerte.
2) **Originalserie –** Dies ist ein Zeitseriendiagramm der realen Daten.
3) **Prognosefehler –** Hierbei handelt es sich um den Unterschied (d.h., um den Fehler) zwischen Prognose und realen Daten.
4) **Saisonunabhängige Prognoseüberlagerung –** Dies ist ein Zeitseriendiagramm der durch die Prognose erstellten Datenwerte, nachdem die Originaldaten saisonunabhängig gemacht wurden.
5) **Saisonunabhängige Originalserie –** Dies ist ein Zeitseriendiagramm der realen Daten, nachdem diese saisonunabhängig gemacht wurden.
6) **Saisonunabhängige Prognosefehler –** Dies sind die Unterschiede (d.h., die Fehler) zwischen der Prognose und den realen Daten, nachdem die Originaldaten saisonunabhängig gemacht wurden.

Prognosebericht (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen zu *Anfang* der Zeitserie kein Problem dar, können aber in der Mitte oder am Ende der Zeitserie nicht mehr gehandhabt werden.
- **Mit Daten verknüpfen –** Wegen der erforderlichen sehr langen Berechnungen werden Prognosen nicht mit den Daten verknüpft. Falls Änderungen an den Originaldaten vorgenommen werden, sollte der Vorgang erneut ausgeführt werden.

Menü *Regression und Klassifizierung*

Über die in diesem Menü enthaltenden Befehle können Regressions- und Klassifizierungsanalysen ausgeführt werden. Bei den verfügbaren Regressionsanalysen handelt es sich um **einfache Mehrfachregression, schrittweise Regression, Vorwärtsregression, Rückwärtsregression** und **Blockregression**. Andere, über dieses Menü auszuführende Analysen schließen die **Diskriminanzanalyse** und die **logische Regression** mit ein.

Befehl Regression

Führt Regressionsanalysen an einem Satz von Variablen aus

Die Regressionsanalyse unterstützt verschiedene Regressionsmodelle, einschließlich Pseudovariablen-Regression (Regression mit kategorischen Variablen), Polynom-Regression und Regression mit Interaktionen zwischen Variablen. Mit dem optionalen Regressionsassistenten können Sie diese Modelle definieren, falls Sie diese verwenden möchten.

Für jedes dieser Modelle sind verschiedene Methoden zur Auswahl der Variablen verfügbar, die in die Gleichung aufgenommen werden sollen. Diese werden als „Regressionstypen“ bezeichnet und dabei handelt es sich um die Methoden Mehrfache Regression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression und Blockregression. Berichte über die einzelnen Regressionen enthalten folgende Einzelheiten: eine ANOVA-Tabelle und eine Tabelle der geschätzten Regressionskoeffizienten nebst zugehörigem Standardfehler, t-Werten, p-Werten und auch entsprechende 95-prozentige Vertrauensbereiche.

Auch haben Sie die Möglichkeit, zwei neue Variablen zu erstellen, d. h., für angepasste und Residualwerte, sowie eine Anzahl von diagnostischen Punktdiagrammen und anderen weiter unten beschriebenen Optionen zu erstellen.

Regressions-assistent

Der Regressionsassistent bietet eine einfache Möglichkeit zum Erstellen von Regressionsmodellen wie Pseudovariablen-Regression oder Polynom-Regression. Für diese Modelle müssen Variablen-Transformationen in der Analyse miteingeschlossen werden („abgeleitete Variablen"); der Assistent erleichtert das Miteinschließen der Transformationen.

Der Assistent führt Sie durch vier Schritte:

Schritt 1: Modellauswahl – Hier geben Sie den Typ des Modells an, das Sie erstellen möchten. Es sind folgende Optionen verfügbar:

- **Standardregressionsmodell –** Mehrfachregressionsmodell.
- **Pseudovariablen-Regressionsmodell –** Mehrfachregressionsmodell, das kategorieunabhängige Variablen enthält. Die 0/1-Codierung für die Kategorien wird automatisch durchgeführt.
- **Polynom-Regressionsmodell –** Mehrfachregressionsmodell, das die Güte der unabhängigen Variablen enthält. Es können auch alle möglichen Interaktionen zwischen den ursprünglichen Variablen berücksichtigt werden.
- **Regressionsmodell mit Interaktionen zwischen Variablen –** Mehrfachregressionsmodell, das alle möglichen Interaktionen zwischen unabhängigen Variablen enthält.
- **Regressionsmodell mit nichtlinearen Variablentransformationen –** Regressionsmodell, bei dem einige nichtlineare Transformationen auf die Variablen angewandt werden. Es sind folgende Optionen verfügbar:

Modell	Regressionsgleichung
Exponentiell	$log(y) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p$
Reziprok	$1/y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p$
Logarithmisch	$y = b_0 + b_1log(x_1) + b_2log(x_2) + \ldots + b_plog(x_p)$
Güte	$log(y) = b_0 + b_1log(x_1) + b_2log(x_2) + \ldots + b_plog(x_p)$

Schritt 2: Auswahl von unabhängigen Variablen – Hier wählen Sie eine oder mehrere unabhängige Variablen und etwaige erwünschte Variablentransformationen aus.

Schritt 3: **Auswahl der abhängigen Variable –** Hier wählen Sie die abhängige Variable aus. Bei einigen Modelltypen wird auch eine Transformation der abhängigen Variable ausgewählt.

Schritt 4: Auswahl der Auswahlmethode – Hier wählen Sie den „Regressionstyp" aus – Mehrfachregression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression oder Blockregression. Wenn der Assistent fertig ist, können die Einstellungen im Dialogfeld Regression verändert werden.

Verfügbare Regressions-typen (Methoden der Variablenauswahl)

Bei den verfügbaren Regressionstypen (Methoden zur Variablenauswahl) handelt es sich um **Mehrfachregression**, **Schrittweise Regression**, **Vorwärtsregression**, **Rückwärtsregression** und **Blockregression**.

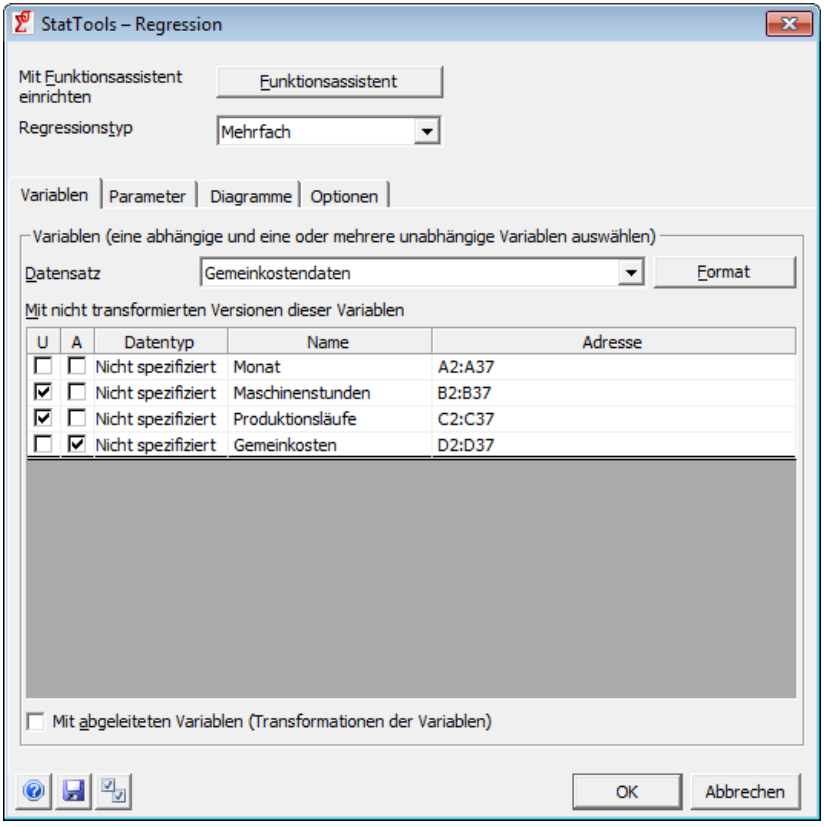
Durch die Mehrfachregression wird sofort eine komplette Gleichung erstellt, und zwar unter Verwendung der ausgewählten unabhängigen Variablen. (Das umfasst den Vorgang, der gewöhnlich als „einfache" Regression bezeichnet wird, bei dem es eine einzige unabhängige Variable gibt.) Bei den anderen Regressionsverfahren können die Variablen (oder Variablenblöcke) sequenziell der Gleichung hinzugefügt oder aus der Gleichung entfernt werden. Insbesondere können Variablen bei der schrittweisen Regression eine nach der anderen hinzugefügt werden. Dabei ist die als nächste hinzuzufügende immer die Variable, die am stärksten mit dem unerklärten Teil der abhängigen Variable korreliert ist.

Bei der schrittweisen Option können Variablen nach der Hinzufügung außerdem auch wieder entfernt werden, wenn sie keinen erheblichen Beitrag mehr leisten können. Die Vorwärtsregression entspricht praktisch der schrittweisen Regression, mit dem Unterschied, dass die Variablen bei der Vorwärtsregression nach Hinzufügung nicht wieder entfernt werden können. Bei der Rückwärtsregression sind zu Beginn alle möglichen unabhängigen Variablen in der Gleichung enthalten und werden dann einzeln gelöscht, sobald sie keinen bedeutsamen Beitrag mehr leisten.

Die Blockregression ermöglicht es schließlich, dass unabhängige Variablen in Form von *Blöcken* hinzugefügt oder nicht hinzugefügt werden, und zwar in einer bestimmten Reihenfolge. Wenn ein Block nicht signifikant ist und deshalb nicht hinzugefügt wird, werden auch keine späteren Blöcke mehr berücksichtigt.

Zum Definieren dieser Analysen wird das Dialogfeld **Regression** verwendet:

Dialogfeld Regression

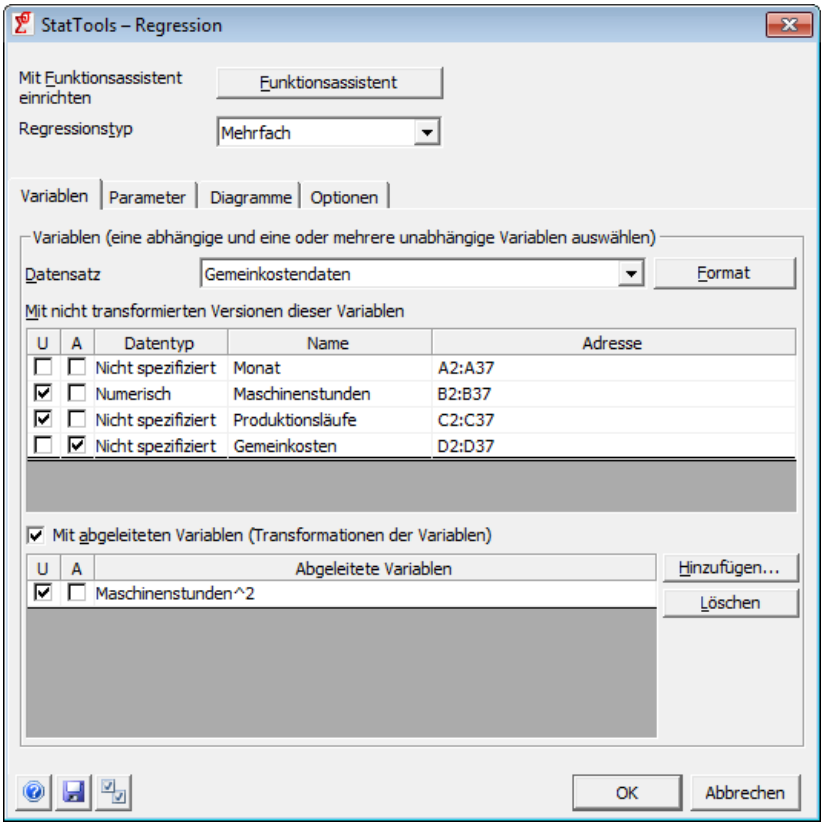


Im Dialogfeld Regression stehen folgende Optionen zur Verfügung:

- **Mit Funktionsassistent einrichten –** Mit dieser Option werden Sie in einer Folge von Dialogfeldern durch den Vorgang zum Erstellen verschiedener Regressionsmodelle geführt.
- **Regressionstyp –** Hier wählen Sie die Methode zur Variablenauswahl aus: Mehrfachregression, Schrittweise Regression, Vorwärtsregression, Rückwärtsregression oder Blockregression.

- **Unabhängige Variablen festlegen –** Wenn Sie diese Option aktivieren, enthält der Variablen-Auswahlraster eine zusätzliche Spalte F. Die Spalte F ermöglicht Ihnen die Auswahl „fixer“ Variablen. Das sind Variablen, die beim Variablen-Auswahlvorgang niemals aus der Gleichung ausgeschlossen werden.

Dialogfeld Regression – Registerkarte Variablen

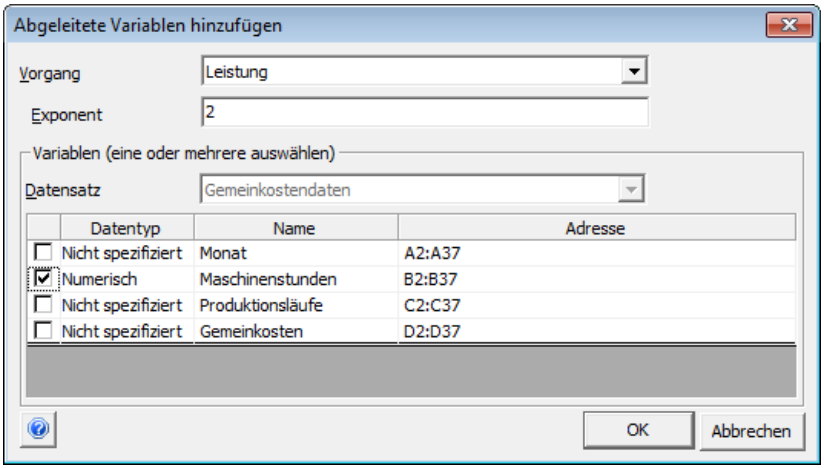


Für diese Analysen müssen eine abhängige Variable (D) und eine oder mehrere unabhängige Variablen (I) ausgewählt werden, aber nicht im Fall der Blockregression. Bei der Blockregression ist es erforderlich, eine abhängige Variable (D) und ein bis sieben Blöcke (B1 bis B7) unabhängiger Variablen auszuwählen.

Da kategorische Variablen transformiert werden müssen, bevor sie in einer Regressionsanalyse verwendet werden können, müssen ausgewählte Variablen entweder als „kategorische“ oder „numerische“" Variablen bezeichnet werden. Bei dem ausgewählten Datensatz muss es sich immer um einen entstapelten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen. Die Option für abgeleitete Variablen ist jedoch bei mehreren Datensätzen nicht verfügbar.

Wenn das Kontrollkästchen **Mit abgeleiteten Variablen** aktiviert ist, können Transformationen von Variablen – wie das Quadrat einer Variable – der Analyse hinzugefügt werden. Wenn Sie auf die Schaltfläche **Hinzufügen** klicken, wird das Dialogfeld **Abgeleitete Variablen hinzufügen** angezeigt.

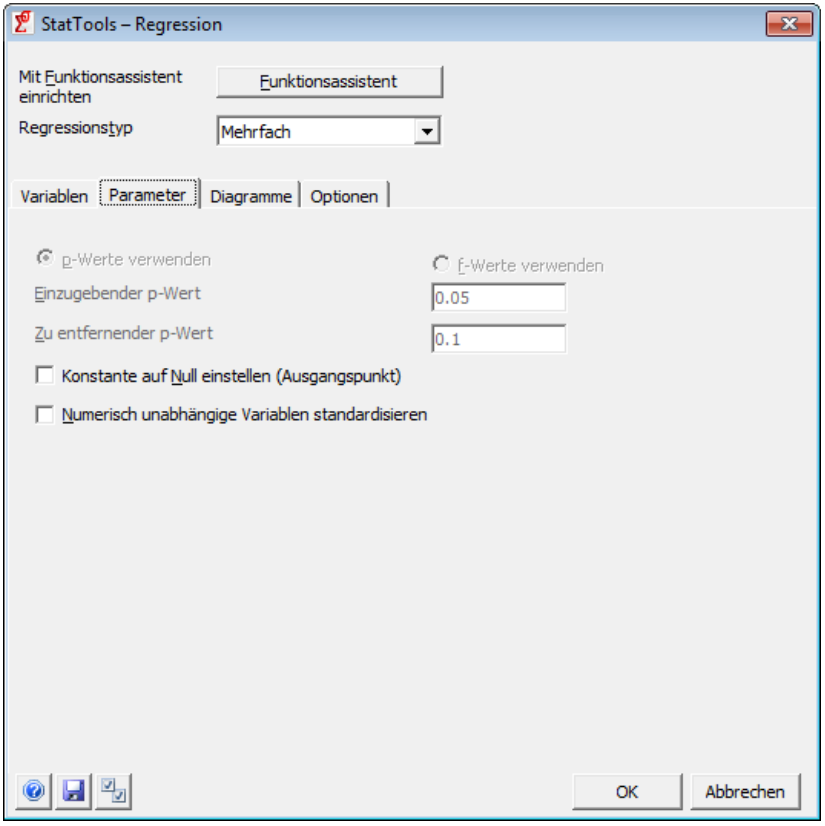
Dialogfeld Abgeleitete Variablen hinzufügen



Das Dialogfeld **Abgeleitete Variablen hinzufügen** bietet eine schnelle und einfache Methode zum Erstellen neuer Variablen, die Transformationen der ursprünglichen Variablen sind. Es sind folgende Optionen verfügbar:

- **Güte –** Wählen Sie diese Option aus, um Gütetransformationen auf die Variablen anzuwenden.
- **Logarithmus –** Wählen Sie diese Option aus, um logarithmische Transformationen auf die Variablen anzuwenden.
- **Interaktion ohne Kategorievariable –** Wählen Sie diese Option aus, um einen Interaktionsbegriff hinzuzufügen, der das Produkt von zwei oder mehr numerischen Variablen ist.
- **Interaktion mit Kategorievariable –** Wählen Sie diese Option aus, um einen Interaktionsbegriff auf Basis von einer oder mehreren numerischen Variablen und einer kategorischen Variable hinzuzufügen.

Dialogfeld Regression – Registerkarte Parameter

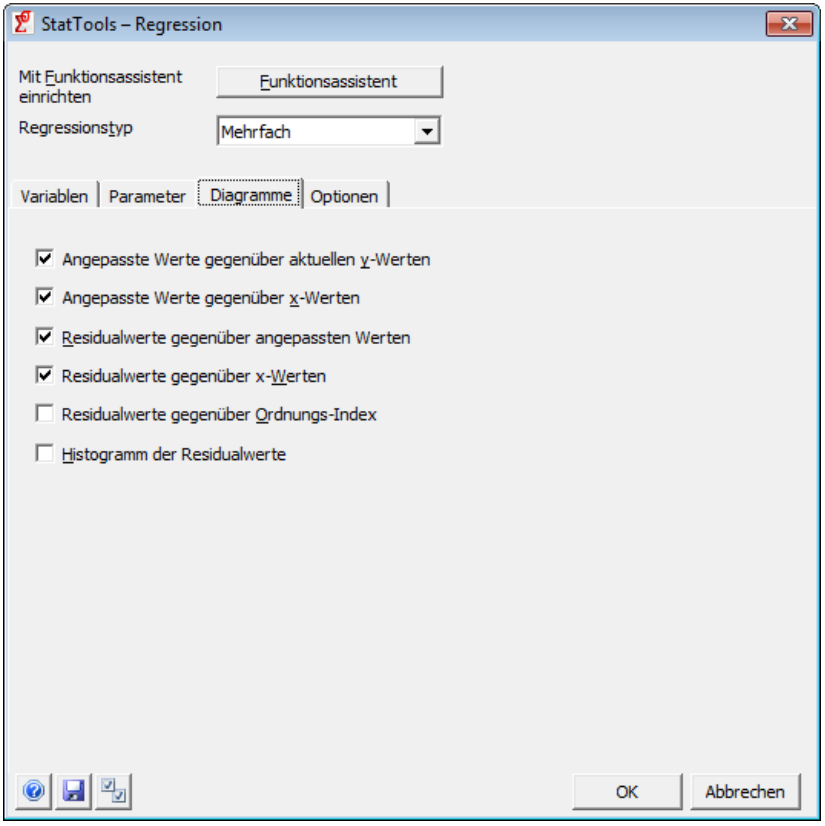


Für Parameter sind folgende Optionen verfügbar:

- **p-Werte verwenden** – Wirkt sich darauf aus, wie die Variablen der Regressionsgleichung hinzugefügt oder daraus entfernt werden. Wenn diese Option ausgewählt ist, können Sie je nach verwendeter Regressionsmethode angeben, dass ein p-Wert eingegeben oder entfernt werden soll. Je geringer der p-Wert ist, desto wichtiger muss eine Variable sein, um zur Regressionsgleichung hinzugefügt oder daraus entfernt werden zu können. Die gezeigten Standardwerte sind gewöhnlich akzeptabel. Sie müssen jedoch zwei Dinge berücksichtigen. 1) Der einzugebende p-Wert darf nicht größer als der zu entfernende p-Wert sein. 2) Um Variablen leichter hinzufügen (und nicht so leicht entfernen) zu können, sollten größere p-Werte verwendet werden. Um Variablen dagegen schwieriger hinzufügen (und leichter entfernen) zu können, sollten kleinere p-Werte verwendet werden. p-Werte liegen gewöhnlich im Bereich von 0,01 bis 0,1.

- **f-Werte verwenden** – Wirkt sich (wie p-Werte) darauf aus, wie die Variablen der Regressionsgleichung hinzugefügt oder daraus entfernt werden. Diese Option ermöglicht Ihnen, je nach verwendeter Regressionsmethode anzugeben, dass ein f-Wert eingegeben oder entfernt werden soll. f-Werte liegen gewöhnlich im Bereich von 2,5 bis 4.
- **Konstante auf Null einstellen** – Erzwingt, dass die Regression durch den Ausgangspunkt verläuft.
- **Numerisch unabhängige Variablen standardisieren** – Der Mittelwert wird von jeder Variable abgezogen und das Ergebnis dann durch die Standardabweichung dividiert. Das ist manchmal von Vorteil, wenn die unabhängigen Variablen sehr unterschiedliche Größenordnungen haben.

Dialogfeld Regression – Registerkarte Diagramme

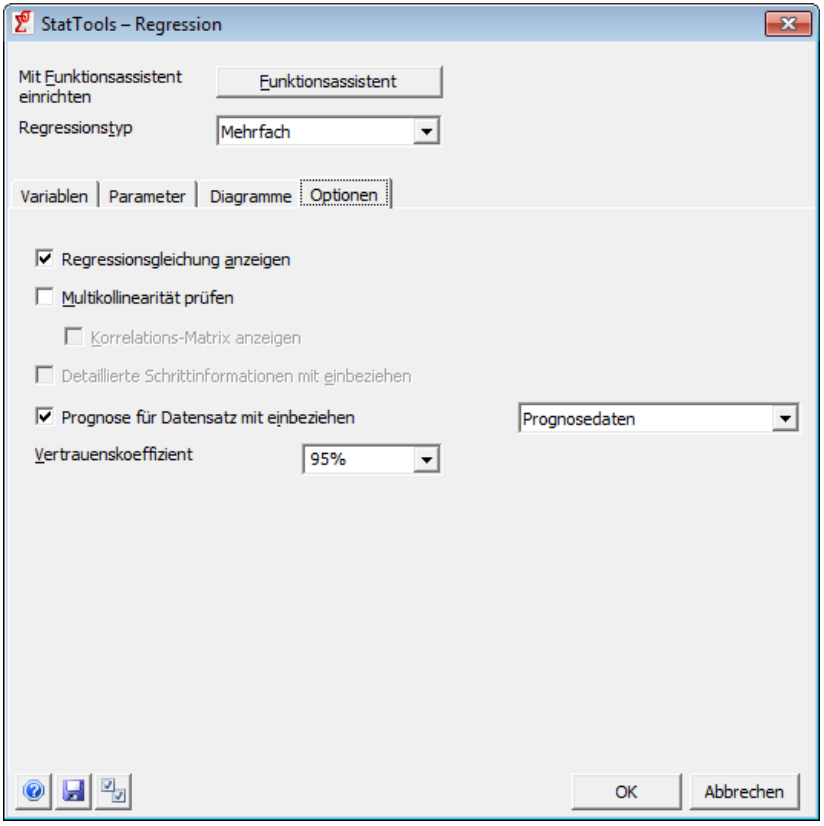


Durch StatTools können mehrere optionale Punktdiagramme erstellt werden, wie im Dialogfeld unter Diagramme zu sehen ist. Dabei handelt es sich u. a. um:

- **Angepasste Werte gegenüber aktuellen y-Werten**
- **Angepasste Werte gegenüber x-Werten**
- **Residualwerte gegenüber angepassten Werten**
- **Residualwerte gegenüber x-Werten**
- **Residualwerte gegenüber Ordnungs-Index**
- **Histogramm der Residualwerte**

Diese Diagramme werden in der Regel bei der Residualwert-Analyse verwendet, um zu überprüfen, ob den Regressionsannahmen entsprochen wird.

Dialogfeld Regression – Registerkarte Optionen



Unter den erweiterten Optionen für die Regressionsanalyse ist Folgendes zu finden:

- **Regressionsgleichung anzeigen** – Dies verdeutlicht die Regressionsgleichungen für den Bericht. Es ist nützlich, wenn eine oder mehrere kategorische Variablen berücksichtigt werden. In diesem Fall wird für jede Kombination der Kategorien eine eigene Gleichung gezeigt.
- **Multikollinearität prüfen** – Das umfasst den Varianz-Hochrechnungsfaktor (VIF) der einzelnen unabhängigen Variablen im Bericht. Große VIF-Werte weisen auf Multikollinearität hin. Die Korrelations-Matrix kann auch berücksichtigt werden.
- **Detaillierte Schrittinformationen mit einbeziehen** – Dies zeigt R-Quadrat- und Standardfehler-Statistiken bei jedem Zwischenschritt der Regression (bei Verwendung eines beliebigen Vorgangs außer Mehrfachregression).

- **Prognose für Datensatz mit einbeziehen** – Das erstellt vorausgesagte Werte für die abhängige Variable in einem zweiten Datensatz für die unabhängigen Variablen-Werte. Dieser Prognosedatensatz muss die gleichen Variablen-Namen wie der Originaldatensatz enthalten, der durch die Regression analysiert wird. Typischerweise sind im Prognosedatensatz Gruppen von Werten für unabhängige Variablen enthalten, für die Sie den Wert für die abhängige Variable voraussagen möchten. Um diese Prognosen vorzunehmen, wird die für den ersten Datensatz berechnete Regressionsgleichung verwendet. Die vorausgesagten Werte für die abhängige Variable werden dann direkt in den Prognosedatensatz eingegeben, indem diese Werte einfach in die Spalte (oder Zeile) für die abhängige Variable eingefüllt werden. Die Prognose-Intervalle und Vertrauensbereiche (für die Mittelwerte) werden auch berücksichtigt.

Regressions-bericht

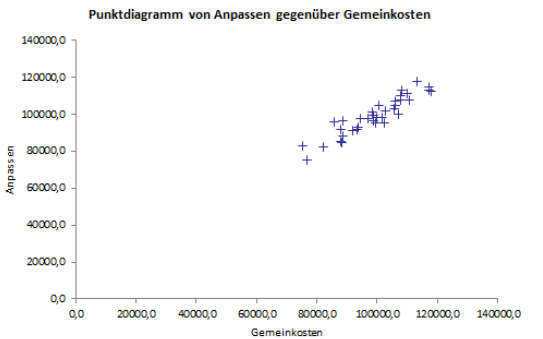
StatTools-Bericht

Analyse: Regression
Ausgeführt durch: Palisade
Datum: Donnerstag, 18. Juni 2015
Aktualisierung: Statisch
Variable: Gemeinkosten

Mehrfachregression für Gemeinkosten *Übersicht*	Mehrere R	r-Quadrat	Berichtigtes r-Quadrat	Schätzungs-Std. Fehler	Zeilen ignoriert
	0,9308	0,8664	0,8583	4108,99309	0

ANOVA-Tabelle	Freiheits-grade	Quadrata-Summe	Quadrata-Mittelwert	F	p-Wert
Erklärt	2	3614020661	1807010330	107,0261279	< 0.0001
Unerklärt	33	557166199,1	16883824,22		

Regressionstabelle	Koeffizient	Standard-fehler	t-Wert	p-Wert	Vertrauensintervall95% Unterer	Oberer
Konstante	3996,678209	6603,650932	0,605222512	0,5492	-9438,550632	17431,90705
Maschinenstunden	43,53639812	3,5894837	12,12887472	< 0.0001	36,23353862	50,83925761
Produktionsläufe	883,6179252	82,25140753	10,74289124	< 0.0001	716,2761784	1050,959672



Datensatz mit Prognosen

Wenn die Option Prognose für Datensatz mit einbeziehen aktiviert ist, werden Prognosen rechts des ausgewählten Datensatzes berücksichtigt. Der Bericht enthält eine Spalte für die Prognosen, zwei Spalten für das Prognosen-Intervall und zwei Spalten für den Vertrauensbereich.

Mitarbeiter	Alter	Geschlecht	BildgStufe	JahreErf	Gehalt	PredictionLowerLimit95	PredictionUpperLimit95	ConfidenceLowerLimit95	ConfidenceUpperLimit95
209	37	Weiblich	3	7	37896,7081	23322,3185	52471,0977	36302,2145	39491,2017
210	49	Männlich	2	10	40207,5287	25376,6608	55038,3966	37031,9626	43383,0947
211	55	Männlich	3	16	54078,3000	39381,5777	68775,0224	51603,7849	56552,8151

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Wenn in einer Zeile irgendwelche Werte für eine ausgewählte Variable fehlen, wird diese Zeile einfach ignoriert.

Mit Daten verknüpfen – Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

Befehl *Logistische Regression*

Führt eine logische Regression an einem Satz von Variablen aus

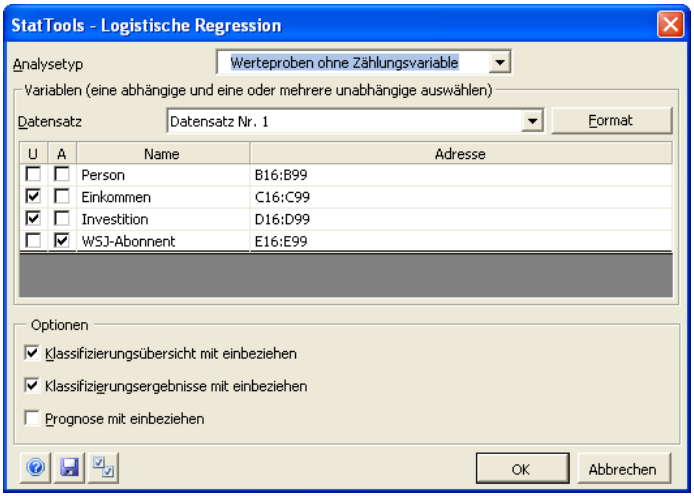
Mithilfe dieses Befehls kann eine logische Regressionsanalyse eines Variablensatzes ausgeführt werden. Dies ist im Wesentlichen eine nicht lineare Art von Regressionsanalyse, wobei dann die Reaktionsvariable binär (d.h., 0 oder 1) ist. Es sollte eine 0-1-Reaktionsvariable vorhanden sein, durch die angegeben wird, ob die einzelnen Beobachtungen jeweils als „Erfolg“ oder „Misserfolg“ zu werten sind. Auch sollten eine oder mehrere zusätzliche Variablen verfügbar sein, durch die die Erfolgswahrscheinlichkeit geschätzt werden kann.

Als zweite Option für die logistische Regression ist eine Zählungsvariable vorhanden, durch die die Anzahl der „Versuche“ angegeben wird, die bei jeder Kombination von zusätzlichen Variablen beobachtet wurden. Auch sollte durch die Reaktionsvariable angezeigt werden, wie viele Versuche erfolgreich waren. Das Ergebnis der logischen Regression ist eine Regressionsgleichung, die ungefähr einer regulären Mehrfachregressionsgleichung entspricht. Wie nachstehend erklärt, muss diese jedoch etwas anders ausgewertet werden.

In StatTools ist der logische Regressionsvorgang auf Optimierung angewiesen, um nach der Regressionsgleichung zu suchen. Für diese Optimierung muss ein komplizierter, nicht linearer Algorithmus verwendet werden, wodurch dieser Vorgang (je nach Geschwindigkeit des PC) einige Zeit in Anspruch nehmen kann.

Dialogfeld Logistische Regression

Zum Definieren dieser Analyse wird das Dialogfeld **Logistische Regression** verwendet:



Für diese Analyse müssen eine abhängige oder Reaktionsvariable (D) und eine oder mehrere unabhängige Variablen (I) ausgewählt werden. Die Daten müssen

1) gestapelt sein und es muss auch eine 0-1-Reaktionsvariable vorhanden sein, aus der hervorgeht, ob die einzelnen Beobachtungen als „Erfolg“ oder „Misserfolg“ gewertet werden können. Dies wird hier **Werteproben ohne Zählungsvariable** genannt.

2) mit einer Zählungsvariable und einer Ganzzahl-Reaktionsvariable versehen sein. Dies wird hier **Übersicht über Werteproben mit Zählungsvariable** genannt.

Die Variablen können aus verschiedenen Datensätzen stammen. Falls der Analysetyp auf **Übersicht über Werteproben mit Zählungsvariable** eingestellt ist, muss eine zusätzliche Zählungsvariable ausgewählt werden.

Im Dialogfeld **Logistische Regression** stehen folgende Optionen zur Verfügung:

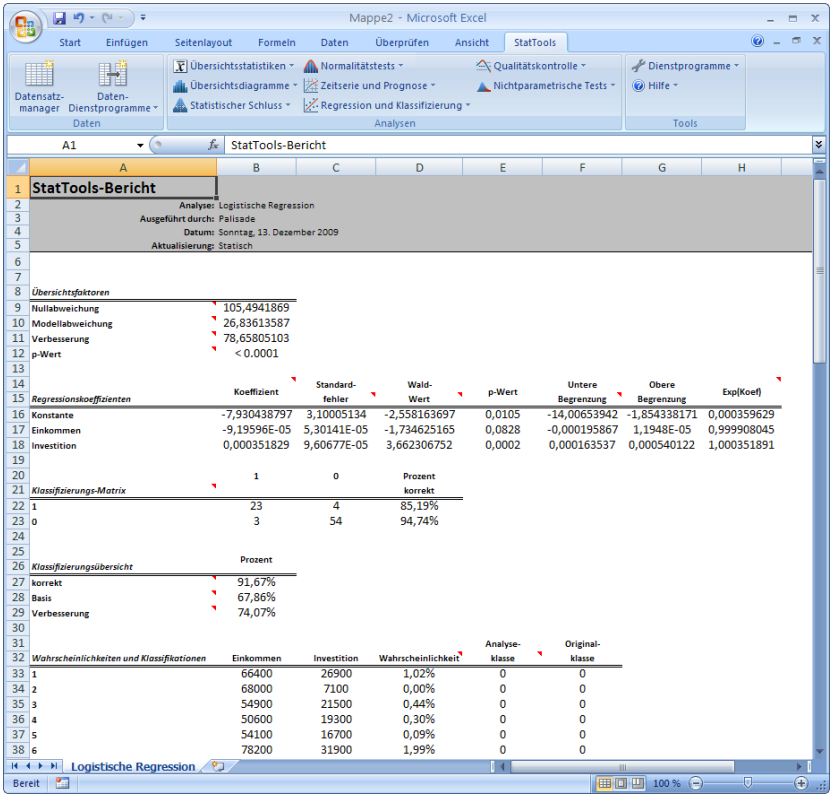
- **Analysetyp** – wählt den Typ der auszuführenden logistischen Regression aus: **Werteproben ohne Zählungsvariable** oder **Übersicht über Werteproben mit Zählungsvariable**.

Bei **Werteproben ohne Zählungsvariable** ist eine 0-1-Reaktionsvariable vorhanden, durch die angegeben wird, ob die einzelnen Beobachtungen jeweils als „Erfolg“ oder „Misserfolg“ zu werten sind. Auch sollten eine oder mehrere zusätzliche Variablen verfügbar sein, durch die die Erfolgswahrscheinlichkeit geschätzt werden kann.

Bei **Übersicht über Werteproben mit Zählungsvariable** ist eine Zählungsvariable vorhanden, durch die die Anzahl der „Versuche“ angegeben wird, die bei jeder Kombination von zusätzlichen Variablen beobachtet wurden. Auch sollte durch die Reaktionsvariable angezeigt werden, wie viele Versuche erfolgreich waren. Bei diesem Analysetyp ist im Variablen-Selektor eine separate Spalte **C** zu sehen, wodurch die Zählungsvariable ausgewählt werden kann.

- **Klassifizierungsübersicht mit einbeziehen** – bezieht sich auf den Regressionsbericht.
- **Klassifizierungsergebnisse mit einbeziehen** – bezieht sich auf den Regressionsbericht.
- **Prognose mit einbeziehen** – bedeutet, dass die vorausgesagten Werte für die abhängige Variable in einem zweiten Datensatz für die unabhängigen Variablen-Werte generiert werden. Dieser **Prognosedatensatz** muss die gleichen Variablen-Namen wie der Originaldatensatz enthalten, der durch die Regression analysiert wird. Typischerweise sind im Prognosedatensatz Gruppen von Werten für unabhängige Variablen enthalten, für die Sie den Wert für die abhängige Variable voraussagen möchten. Um diese Prognosen vorzunehmen, wird die für den ersten Datensatz berechnete Regressionsgleichung verwendet. Die vorausgesagten Werte für die abhängige Variable werden dann direkt in den Prognosedatensatz eingegeben, indem diese Werte einfach in die Spalte (oder Zeile) für die abhängige Variable eingefüllt werden.

Bericht über logistische Regression



Der vorstehende Bericht enthält die Originaldaten plus die Klassifizierungsdaten. Die vorausgesagten Klassifizierungen in Spalte E sind davon abhängig, ob die geschätzten Erfolgswahrscheinlichkeiten (in Spalte D) oberhalb oder unterhalb des Trennwerts von 0,5 (oder 50 %) liegen. In dem Bericht sind Übersichtsstatistik für die Regression (ähnlich der R-Quadrat-Statistik für Mehrfachregression), detaillierte Informationen über die Regressionsgleichung und die Übersicht über die Klassifizierungsergebnisse zu sehen. (Die Zellen enthalten viele hilfreiche Kommentare, um das Auswerten der Ergebnisse zu erleichtern.) Wie wir sehen, sind in diesem Beispiel 90,5 % der Beobachtungen richtig klassifiziert. Es ist natürlich unser Bestreben, hier einen möglichst hohen Prozentsatz zu erhalten. Gewöhnlich können die Ergebnisse nur durch zusätzliche oder bessere Variablen noch verbessert werden. Die Werte in Spalte H **[Exp (Koeff.)]** werden meistens zur Auswertung der Regressionsgleichung verwendet. Aus diesen Werten geht die geschätzte Änderung in den Erfolgschancen hervor, wenn irgendeine der beschreibenden Variablen um eine Einheit erhöht wird.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Wenn in einer Zeile irgendwelche Werte für eine ausgewählte Variable fehlen, wird diese Zeile einfach ignoriert.
- **Mit Daten verknüpften** – Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

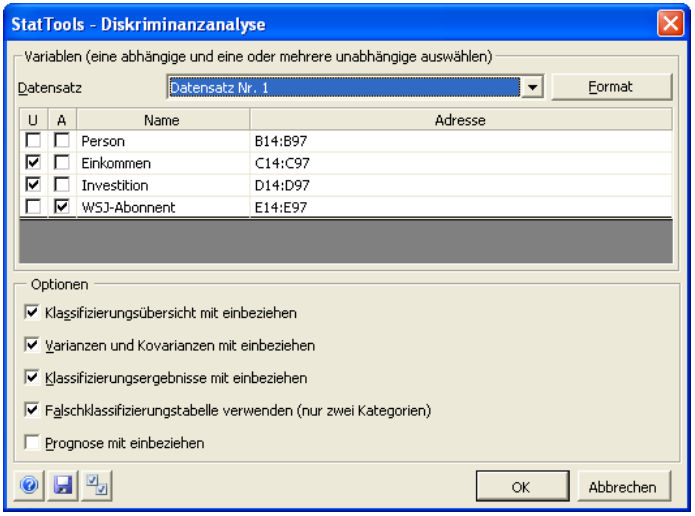
Befehl *Diskriminanzanalyse*

Führt eine Diskriminanzanalyse eines Satzes von Variablen aus

Über diesen Befehl kann ein Datensatz auf Diskriminanz analysiert werden. Bei dieser Analyse ist eine Kategorievariable vorhanden, aus der hervorgeht, in welcher von zwei oder mehr Gruppen sich die betreffende Beobachtung befindet; auch sollten eine oder mehrere beschreibende Variablen verfügbar sein, um die Gruppenmitgliedschaft voraussagen zu können Das kann auf zwei Weisen geschehen. Die erste und allgemeine Methode, die für jede beliebige Anzahl von Gruppen verwendet werden kann, ist das Berechnen der „statistischen Distanz“ der einzelnen Beobachtungen vom Mittelwert jeder Gruppe und das Klassifizieren der Beobachtung entsprechend der geringsten statistischen Distanz. Bei der zweiten Methode, die im Falle von zwei Gruppen verwendet wird, kann eine Diskriminanzfunktion (d.h., ein linearer Ausdruck der beschreibenden Variablen) berechnet und dann jede Beobachtung klassifiziert werden, und zwar genau danach, ob der entsprechende Diskriminanzwert kleiner oder größer als ein bestimmter Trennwert ist. Mithilfe dieser zweiten Methode können Sie außerdem vorherigen Wahrscheinlichkeiten von Gruppenmitgliedschaft und auch die Falschklassifizierungskosten angeben. In solchem Fall entspricht dann der Klassifizierungsvorgang dem Minimieren der erwarteten Falschklassifizierungskosten.

Dialogfeld Diskriminanzanalyse

Zum Definieren dieser Analyse wird das Dialogfeld **Diskriminanzanalyse** verwendet:



Für diese Analyse müssen eine abhängige Variable (D) und eine oder mehrere unabhängige Variablen (I) ausgewählt werden. Auch müssen die Daten in entstapelter Form vorhanden sein. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

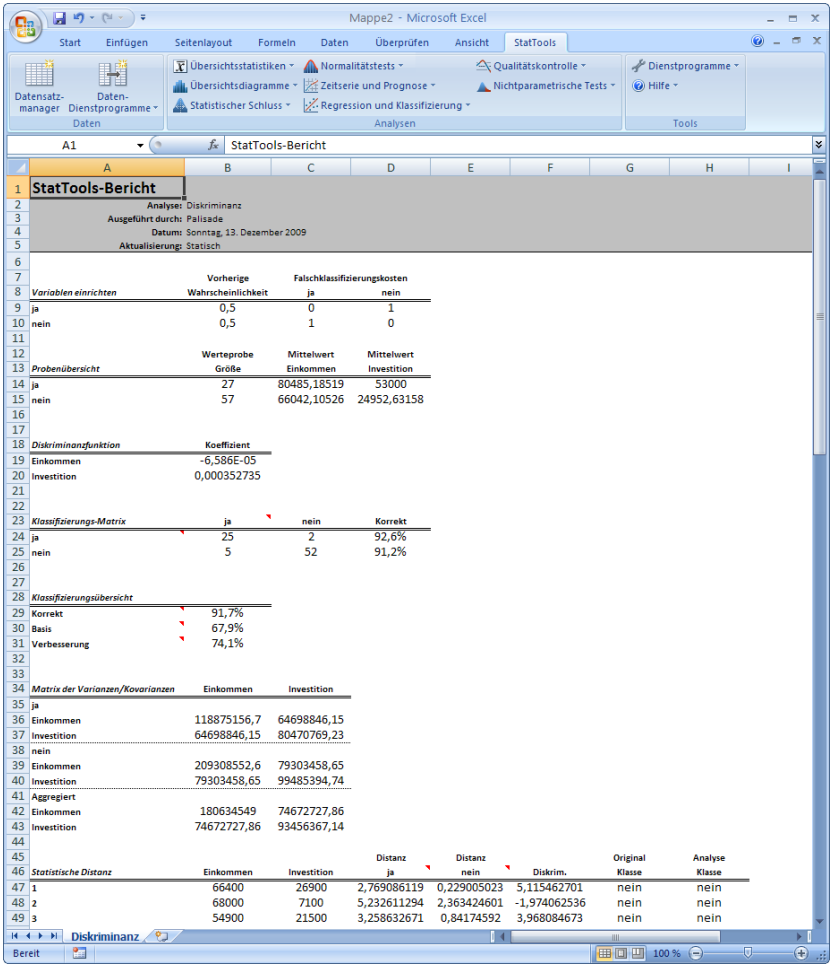
Im Dialogfeld **Diskriminanzanalyse** stehen folgende Optionen zur Verfügung:

- **Klassifizierungsübersicht mit einbeziehen –** bezieht sich auf den Regressionsbericht.
- **Varianzen und Kovarianzen mit einbeziehen –** bezieht sich auf den Regressionsbericht.
- **Klassifizierungsergebnisse mit einbeziehen –** bezieht sich auf den Regressionsbericht.
- **Falschklassifizierungstabelle verwenden –** wird ausgewählt, wenn Sie die vorherigen Wahrscheinlichkeiten oder Falschklassifizierungskosten ändern möchten.
- **Prognose mit einbeziehen –** bedeutet, dass die vorausgesagten Werte für die abhängige Variable in einem zweiten Datensatz für die unabhängigen Variablen-Werte generiert werden. Dieser **Prognosedatensatz** muss die gleichen Variablen-Namen wie der Originaldatensatz enthalten, der durch die Regression analysiert wird. Typischerweise sind im Prognosedatensatz Gruppen von Werten für unabhängige Variablen enthalten, für die Sie den Wert für die abhängige Variable voraussagen möchten. Um diese Prognosen vorzunehmen, wird die für den ersten Datensatz berechnete Regressionsgleichung verwendet. Die vorausgesagten Werte für die abhängige Variable werden dann direkt in den Prognosedatensatz eingegeben, indem diese Werte einfach in die Spalte (oder Zeile) für die abhängige Variable eingefüllt werden.

Dialogfeld Falschklassifizierungskosten

Falls (wie in diesem Beispiel) genau zwei Gruppen für die abhängige Kategorievariable möglich sind und die Option **Falschklassifizierungstabelle verwenden** ausgewählt ist, wird ein Dialogfeld eingeblendet, in dem Sie vorherigen Wahrscheinlichkeiten und/oder Falschklassifizierungskosten angeben können. Nach den Standardeinstellungen ist jede Gruppe gleichermaßen wahrscheinlich und sind auch die Falschklassifizierungskosten gleich, aber Sie können auch andere Einstellungen vornehmen.

Bericht über Diskriminanzanalyse



Dieser Bericht enthält die Originaldaten plus die Klassifizierungsdaten. Die vorausgesagten Klassifizierungen richten sich danach, ob die gezeigten Diskriminanzwerte unterhalb oder oberhalb des betreffenden Trennwerts liegen. Wenn die vorherigen Wahrscheinlichkeiten und Falschklassifizierungskosten auf den Standardwerten gehalten werden, entspricht dieser Vorgang einer Klassifizierung, die auf der kleineren von zwei statistischen Distanzen basiert. Wenn in der Tat mehr als zwei Gruppen vorhanden wären, würden die Diskriminanzwerte nicht in Erscheinung treten, und die Klassifizierung würde dann auf Basis der kleinsten der statistischen Distanzen vorgenommen werden. In dem Bericht ist auch eine beschreibende Statistik in Bezug auf die Gruppen und Koeffizienten der Diskriminanzfunktion zu sehen (aber nur, wenn zwei Gruppen vorhanden sind). Ferner enthält der Bericht die vorherigen Wahrscheinlichkeiten, Falschklassifizierungskosten und den Trennwert für Falschklassifizierung (aber wiederum nur, wenn zwei Gruppen vorhanden sind) sowie auch die Ergebnisübersicht über das Klassifizierungsverfahren (und zwar mit entsprechenden Kommentaren in den Zellen, um die Ergebnisse besser auswerten zu können). Wie wir sehen, sind in diesem Beispiel 89% der Beobachtungen richtig klassifiziert. Es ist natürlich unser Bestreben, hier einen möglichst hohen Prozentsatz zu erhalten. Gewöhnlich können die Ergebnisse nur durch zusätzliche (oder bessere) beschreibende Variablen noch verbessert werden.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Wenn in einer Zeile irgendwelche Werte für eine ausgewählte Variable fehlen, wird diese Zeile einfach ignoriert.
- **Mit Daten verknüpften –** Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

Menü *Qualitätskontrolle*

Die Befehle in diesem Menü haben mit der Analyse von Daten zu tun, die über längere Zeitperioden hinweg erfasst wurden, und zwar mittels Anwendungen für Qualitätskontrolle.

Im Pareto-Diagramm wird die relative Wichtigkeit von kategorisierten Daten angezeigt.

Durch die vier Arten von Steuerdiagrammen können Zeitseriendaten grafisch dargestellt werden, wodurch zu sehen ist, ob ein Vorgang statistisch erfasst wird. Es kann erkannt werden, ob die Daten im Diagramm innerhalb der Steuerbegrenzungen bleiben und wir können auch überprüfen, ob irgendein anderes nicht zufälliges Verhalten (wie z.B. lange Ausführungen oberhalb oder unterhalb der Mittellinie) vorhanden ist.

Befehl *Pareto-Diagramm*

Erstellt ein Pareto-Diagramm für eine kategorisierte Variable

Mithilfe von Pareto-Diagrammen können die wichtigsten Elemente in einer Gruppe von kategorisierten Daten festgestellt und kann auch eine schnelle visuelle Darstellung der relativen Wichtigkeit dieser Elemente gegeben werden. Gewöhnlich werden Pareto-Diagramme für die Qualitätssicherung verwendet, um festzustellen, welche Faktoren am wichtigsten sind (80/20-Regel).

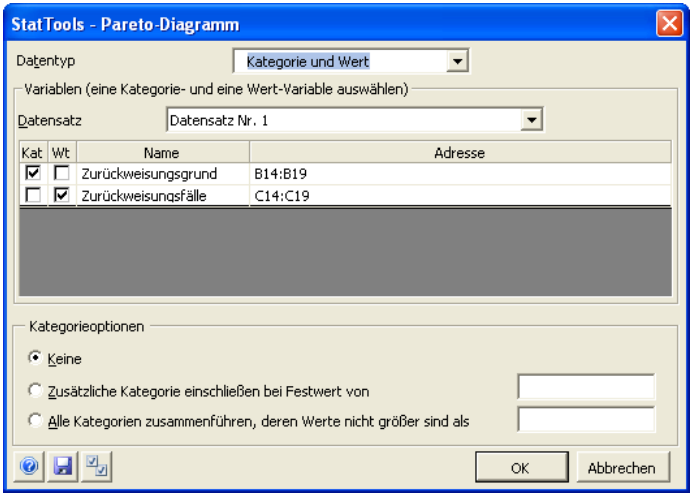
Angenommen, ein Hersteller von Maschinenteilen möchte feststellen, warum die Kunden ein bestimmtes Produkt zurückweisen. Bei Rückgabe der einzelnen Sendungen wird ein Grund (*falsche Größe, fehlerhafte Oberflächenausführung* usw.) eingegeben. Nachdem diese Daten für mehrere Monate erfasst wurden, wird ein Pareto-Diagramm erstellt. Anschließend wird auf die größten Problemquellen eingegangen.

StatTools ermöglicht Ihnen, Pareto-Diagramme auf Basis von Daten in einem von zwei Formaten (*Nur Kategorie* oder *Kategorie und Wert*) zu erstellen. Eine *Nur Kategorie*-Variable enthält in der Regel je einen Eintrag pro Funktion. In dem vorstehenden Beispiel würden die einzelnen Zellen jeweils den Grund dafür enthalten, weshalb eine Sendung von Maschinenteilen zurückgeschickt wurde. Eine Zelle könnte beispielsweise den Grund **fehlerhafte Oberflächenausführung** enthalten und wahrscheinlich würde es viele Zellen geben, die diesen Grund aufweisen. Durch StatTools wird dann gezählt, wie oft jeder Eintrag in der Variable auftritt, und anschließend dann das entsprechende Pareto-Diagramm erstellt. Wenn dagegen *Kategorie und Wert* ausgewählt ist, handelt es sich bei den von Ihnen angegebenen Variablen um die Kategorien und die jeweils entsprechenden Zählungswerte.

In einem Pareto-Diagramm sind die Achsen wie folgt konstruiert:

- Kategorien werden entlang der Horizontalachse platziert
- Häufigkeits- oder Zählungswerte befinden sind entlang der linken Vertikalachse
- und prozentuale Summenwerte sind entlang der rechten Vertikalachse zu sehen

Dialogfeld Pareto-Diagramm

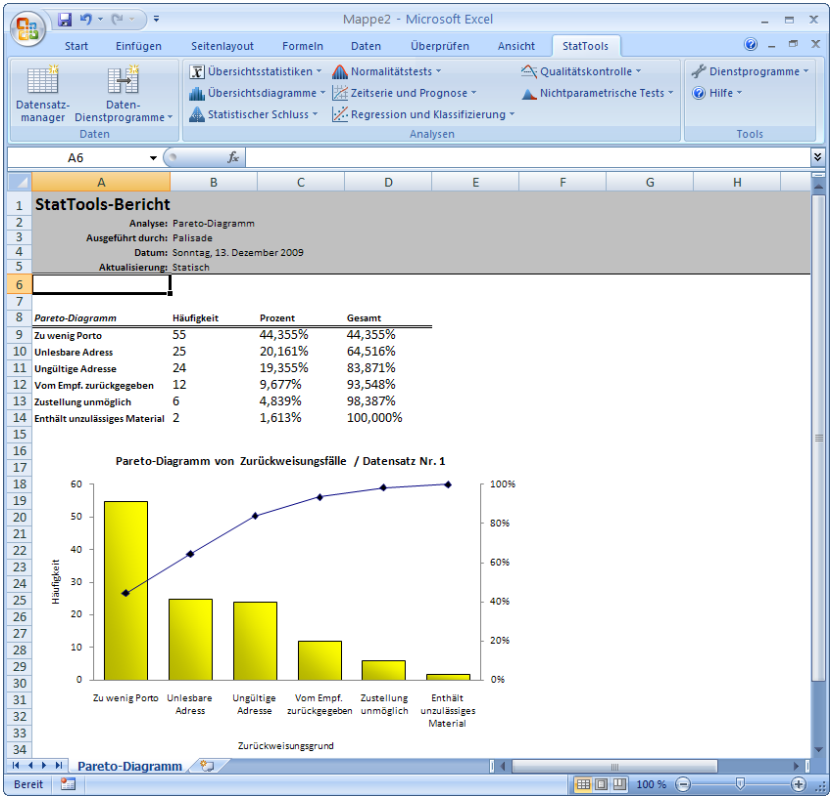


Im Dialogfeld **Pareto-Diagramm** sind u.a. folgende Optionen verfügbar:

- **Datentyp** – wählt den Datentyp aus, der zum Erstellen des Pareto-Diagramms verwendet wird: *Kategorie und Wert* oder *Nur Kategorie*.

Unter **Kategorieoptionen** ist im Dialogfeld **Pareto-Diagramm** Folgendes zu finden:

- **Keine** – bedeutet, dass jede eindeutige Kategorie im Pareto-Diagramm durch einen Balken dargestellt wird.
- **Zusätzliche Kategorie einschließen mit Festwert von** – bedeutet, dass dem Pareto-Diagramm ganz rechts ein Balken mit der Bezeichnung **Sonst**. hinzugefügt wird, und zwar mit einem Häufigkeits- oder Zählungswert, der dem angegebenen Wert entspricht.
- **Alle Kategorien zusammenführen, deren Werte nicht größer sind als** – bedeutet, dass alle Kategorien, deren Häufigkeitswert nicht größer als der angegebene Wert ist, in der Kategorie **Sonst**. zusammengefasst und im Pareto-Diagramm ganz rechts platziert werden.



Pareto-Diagramm	Häufigkeit	Prozent	Gesamt
Zu wenig Porto	55	44,355%	44,355%
Unlesbare Adress	25	20,161%	64,516%
Ungültige Adresse	24	19,355%	83,871%
Vom Empf. zurückgegeben	12	9,677%	93,548%
Zustellung unmöglich	6	4,839%	98,387%
Enthält unzulässiges Material	2	1,613%	100,000%

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Wenn in einer Zeile irgendwelche Werte für eine ausgewählte Variable fehlen, wird diese Zeile einfach ignoriert.
- **Mit Daten verknüpften** – Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

Befehl *x/r-Diagramme*

Erstellt x- und r-Steuerdiagramme für Zeitserien-Variablen

Mithilfe dieser Analyse können x-Balken- und r-Diagramme für Zeitseriendaten erstellt werden. Dabei wird davon ausgegangen, dass die Daten über längere Zeitperioden in kleinen Probenuntermengen erfasst wurden. Es könnten beispielsweise alle 30 Minuten bestimmte Messwerte über die Breiten von vier zufällig gewählten Maschinenteilen erfasst werden. Die Größe der Probenuntermenge ist in diesem Fall 4. Falls Daten aus 50 halbstündigen Perioden erfasst werden, sollten diese Daten in vier nebeneinander befindlichen Spalten und 50 entsprechenden Zeilen angeordnet werden, und zwar mit Überschriften, wie z.B. *Unterm.1* bis *Unterm.4*, über der ersten Datenzeile.

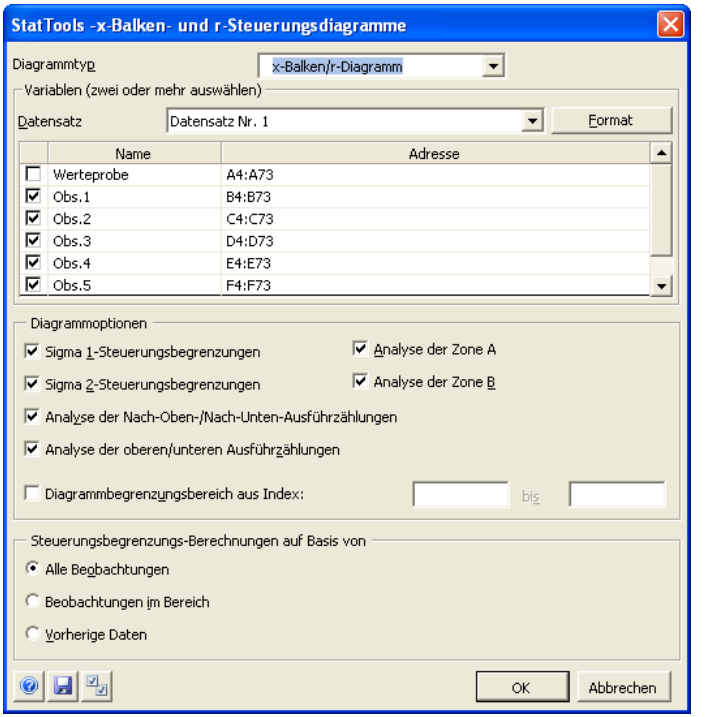
Dieser Vorgang ist dazu da, um zu überprüfen, ob der die Daten generierende Prozess auch statistisch erfasst wird. Zu diesem Zweck wird zuerst im Datensatz ein ***x***-Balken und ein ***r*** für jede Zeile berechnet. ***x***-Balken stellt den Durchschnitt der Beobachtungen in der Zeile dar und ***r*** ist der Beobachtungsbereich (Maximum minus Minimum) in dieser Zeile.

Die ***x***-Balken und ***r***-Bereiche werden um Mittellinien herum als separate Zeitserien-Diagramme grafisch dargestellt. Bei der Mittellinie für das ***x***-Balkendiagramm handelt es sich um den Durchschnitt der ***x***-Balken (mitunter auch ***x-Doppelbalken*** genannt) und die Mittellinie für das ***r***-Diagramm ist der ***r***-Balken (d.h., der Durchschnitt der ***r***-Bereiche). Es kann leicht überprüft werden, ob dieser Prozess richtig erfasst wird, indem Sie nachsehen, ob irgendwelche ***x***-Balken oder ***r***-Bereiche außerhalb der betreffenden oberen (UCL) und unteren (LCL) Steuerbegrenzung liegen, die jeweils ungefähr +/- 3 Standardabweichungen von der Mittellinie entfernt liegt. In den Diagrammen sind diese Steuerbegrenzungen zu sehen, sodass Extremwerte leicht erkannt werden können.

Mithilfe dieses Vorgangs können Sie auch nach anderen möglichen nicht erfassten Verhaltensweisen suchen, z.B. nach 8 oder mehr Punkten in einer Zeile oberhalb oder unterhalb der Mittellinie oder nach 8 oder mehr Punkten in einer Zeile in aufsteigender oder absteigender Richtung. Auch können Sie auf diese Weise mindestens 4 von 5 Punkten in einer Zeile erkennen, die mehr als eine Standardabweichung von der Mittellinie entfernt ist, oder mindestens 2 von 3 Punkten in einer Zeile, die mehr als 2 Standardabweichungen von der Mittellinie entfernt liegt.

Dialogfeld x-Balken- und r-Steuerdiagramme

Zum Definieren dieser Diagramme wird das Dialogfeld **x-Balken- und r-Steuerdiagramme** verwendet:



Für diese Analyse können mehrere Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Unter **Diagrammoptionen** ist in diesem Dialogfeld Folgendes zu finden:

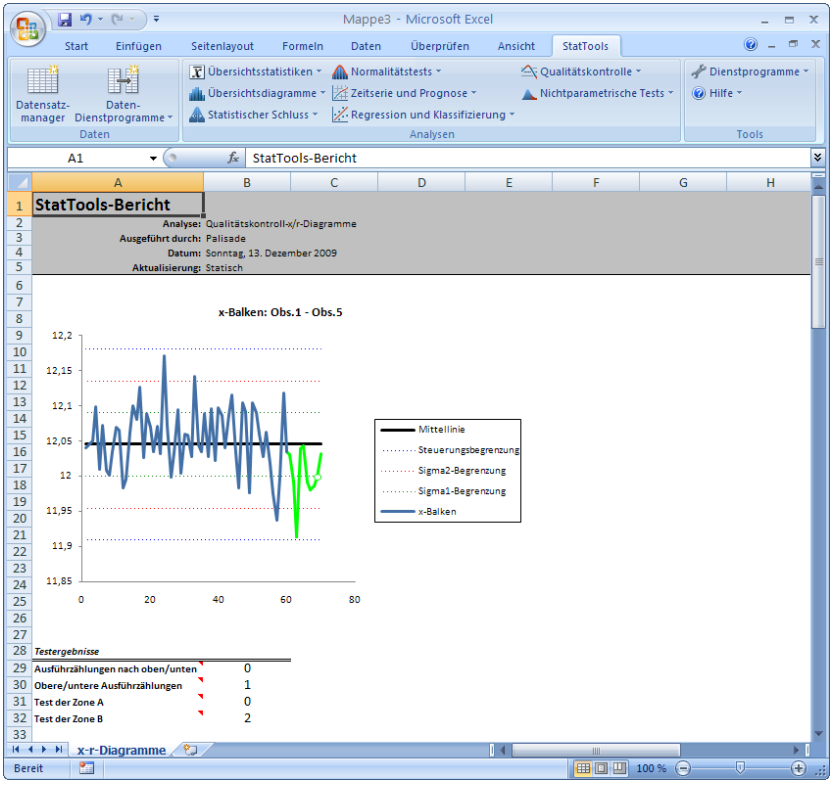
- **Sigma 1-** und **Sigma 2-Steuergrenzen** – fügt bei Sigma 1 oder Sigma 2 von der Mittellinie gewisse Steuergrenzlinien hinzu. Diese zusätzlichen Linien ermöglichen Ihnen, nach anderem nicht erfassten Verhalten (den so genannten „Zonenregeln“) zu suchen.
- **Analyse der Zone A** und **Zone B** – Anzahl der Punkte außerhalb der Zone A (Sigma 2) und Zone B (Sigma 1)
- **Analyse der Nach-Oben-/Nach-Unten-Ausführzählungen** und **Analyse der oberen/unteren Ausführzählungen** – sequenzielle Längenbewegungen nach oben oder unten im Werte von 8 oder mehr

- **Diagrammbegrenzungsbereich aus Index** – begrenzt die Punkte für eine Variable im Diagramm auf einen Datenpunktebereich (d.h., auf den Bereich vom Start- bis zum Endindex)

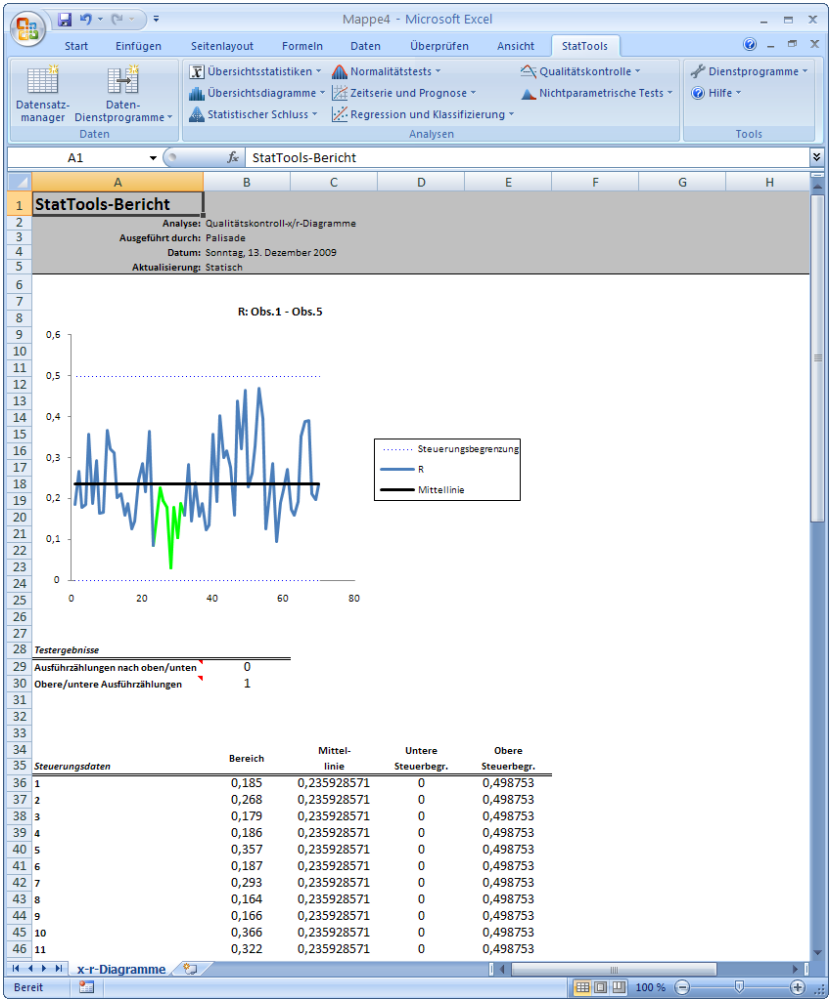
Steuerbegrenzungs-Berechnungen auf Basis von – bestimmt die Daten, die als Basis für die Steuerbegrenzungs-Berechnungen dienen. Folgende Optionen sind hierunter verfügbar:

- **Alle Beobachtungen** – verwendet alle verfügbaren Daten in den Steuerbegrenzungs-Berechnungen
- **Beobachtungen im Bereich** – verwendet in den Steuerbegrenzungs-Berechnungen die Daten zwischen **Startindex** und **Stoppindex**
- **Vorherige Daten** – erstellt Steuerbegrenzungen aus den vorhergehend beobachteten Daten. Dabei brauchen nur **Größe der Probenuntermenge**, **r-Durchschnitt** und **x-Balken-Durchschnitt** der vorhergehend berechneten Daten eingegeben werden.

x-Balkendiagramm (Beispiel)



r-Diagramm (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten können nicht gehandhabt werden.
- **Mit Daten verknüpfen** – Diagramme sind nicht mit den Daten verknüpft.

Befehl *p-Diagramm*

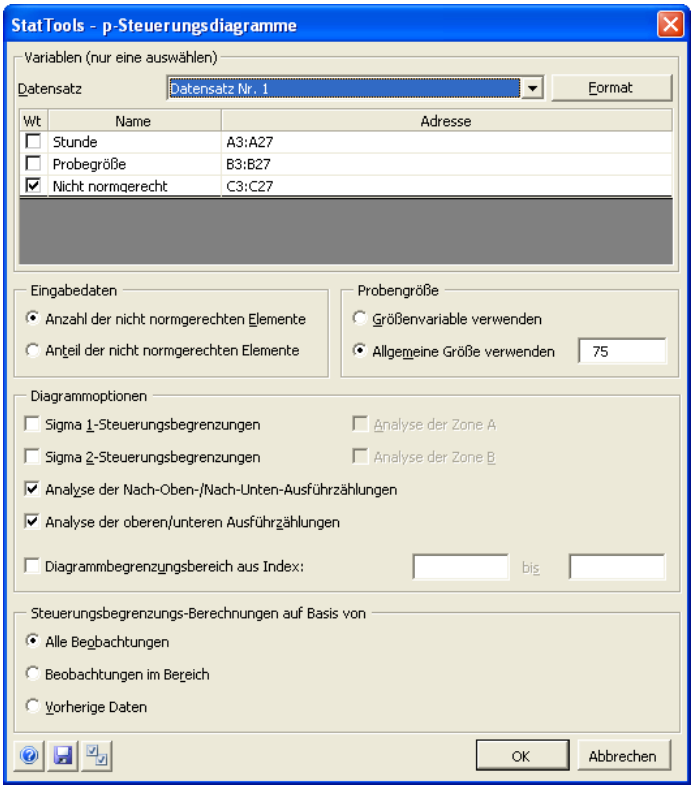
Erstellt *p*-Diagramme für Zeitserien-Variablen

p-Diagramme sind für Attributdaten vorgesehen. Bei Attributdaten wird durch jede Beobachtung die Anzahl der Elemente angezeigt, die nicht den Spezifikationen der betreffenden Elementenprobe entspricht. Durch einen Prozess können möglicherweise halbstündig eine bestimmte Anzahl von Maschinenteilen erstellt werden, von denen einige nicht den Spezifikationen entsprechen. In diesem Fall könnte durch ein *p*-Diagramm halbstündig die Anzahl der Teile grafisch dargestellt werden, die nicht normgerecht sind. Wie immer, soll durch diesen Vorgang herausgefunden werden, ob der Prozess richtig erfasst oder gesteuert wird.

Für dieses Verfahren ist ein Datensatz erforderlich, der mindestens eine Variable enthält, aus der die *Anzahl* der nicht normgerechten Elemente oder Teile in jeder Werteprobe hervorgeht, oder eine Variable, die den *Anteil* der nicht normgerechten Elemente in jeder Werteprobe zeigt. Wahlweise kann auch eine Variable vorhanden sein, aus der die Probengrößen hervorgehen. Wenn solche Variable nicht verwendet wird, müssen Sie eine Probengröße eingeben, von der dann angenommen wird, dass sie für alle Werteproben gleich ist. Wenn jedoch eine Variable für die Probengröße verfügbar ist, kann mit unterschiedlichen Probengrößen gearbeitet werden.

Dialogfeld p-Steuerdiagramme

Zum Definieren dieses Diagramms wird das Dialogfeld **p-Steuerdiagramme** verwendet:



Hier müssen zum Analysieren eine Wertvariable und (wahlweise) eine Variable für die Probengröße ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Unter **Eingabedaten** sind in diesem Dialogfeld folgende Optionen verfägbar:

- **Anzahl der nicht normgerechten Elemente** – legt fest, dass durch die Wertvariable die effektive Anzahl der nicht normgerechten Elemente in der gesamten Werteprobe angegeben wird
- **Anteil der nicht normgerechten Elemente** – legt fest, dass durch die Wertvariable der Anteil der nicht normgerechten Elemente in der Werteprobe angegeben wird

Unter **Probengröße** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Größenvariable verwenden** – legt fest, dass eine Größenvariable verwendet wird, um die Gesamtgröße der einzelnen Werteproben anzugeben
- **Allgemeine Größe verwenden** – legt fest, dass keine Größenvariable verwendet wird, da die einzelnen Werteproben der eingegebenen Größe entsprechen

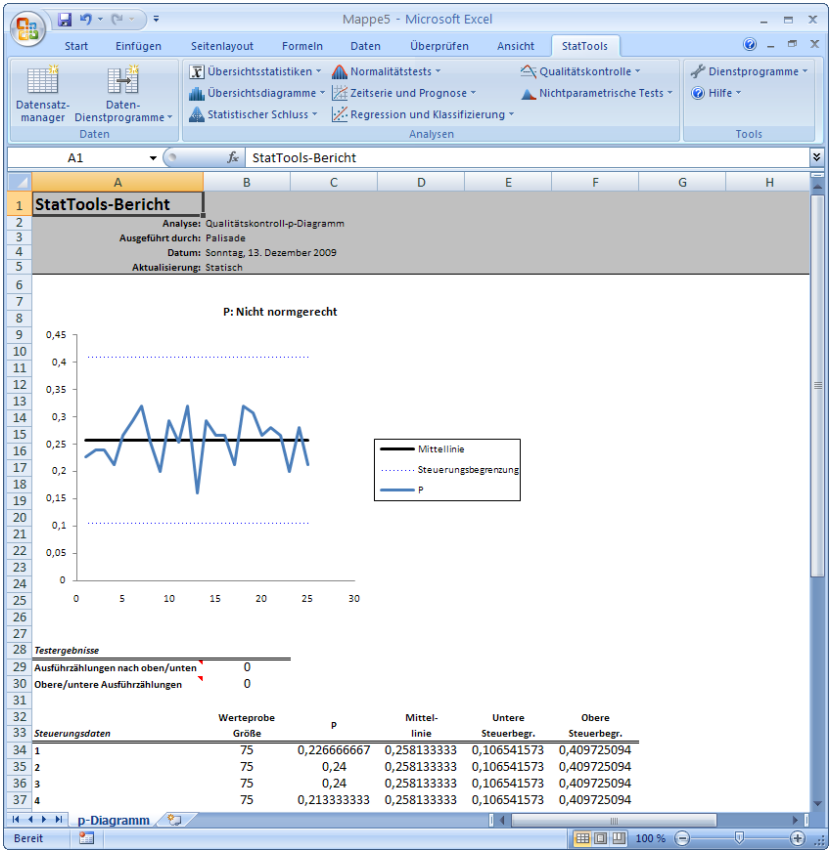
Unter **Diagrammoptionen** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Sigma 1-** und **Sigma 2-Steuergrenzen** – fügt bei Sigma 1 oder Sigma 2 von der Mittellinie gewisse Steuergrenzlinien hinzu. Diese zusätzlichen Linien ermöglichen Ihnen, nach anderem nicht erfassten Verhalten (den so genannten „Zonenregeln") zu suchen.
- **Analyse der Zone A** und **Zone B** – Anzahl der Punkte außerhalb der Zone A (Sigma 2) und Zone B (Sigma 1).
- **Analyse der Nach-Oben-/Nach-Unten-Ausführzählungen** und **Analyse der oberen/unteren Ausführzählungen** – sequenzielle Längenbewegungen nach oben oder unten im Werte von 8 oder mehr
- **Diagrammbegrenzungsbereich aus Index** – begrenzt die Punkte für eine Variable im Diagramm auf einen Datenpunktebereich (d.h., auf den Bereich vom Start- bis zum Endindex)

Steuerbegrenzungs-Berechnungen auf Basis von – bestimmt in diesem Dialogfeld die Daten, die als Basis für die Steuerbegrenzungs-Berechnungen dienen. Folgende Optionen sind verfügbar:

- **Alle Beobachtungen** – verwendet alle verfügbaren Daten in den Steuerbegrenzungs-Berechnungen
- **Beobachtungen im Bereich** – verwendet in den Steuerbegrenzungs-Berechnungen die Daten zwischen **Startindex** und **Stoppindex**
- **Vorherige Daten** – erstellt Steuerbegrenzungen aus den vorhergehend beobachteten Daten. Zu diesem Zweck brauchen nur **Größe der Probenuntermenge** und **p-Durchschnitt** der vorhergehend berechneten Daten eingegeben werden.

p-Diagramm (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten können nicht gehandhabt werden.
- **Mit Daten verknüpfen –** Diagramme sind nicht mit den Daten verknüpft.

Befehl *c-Diagramm*

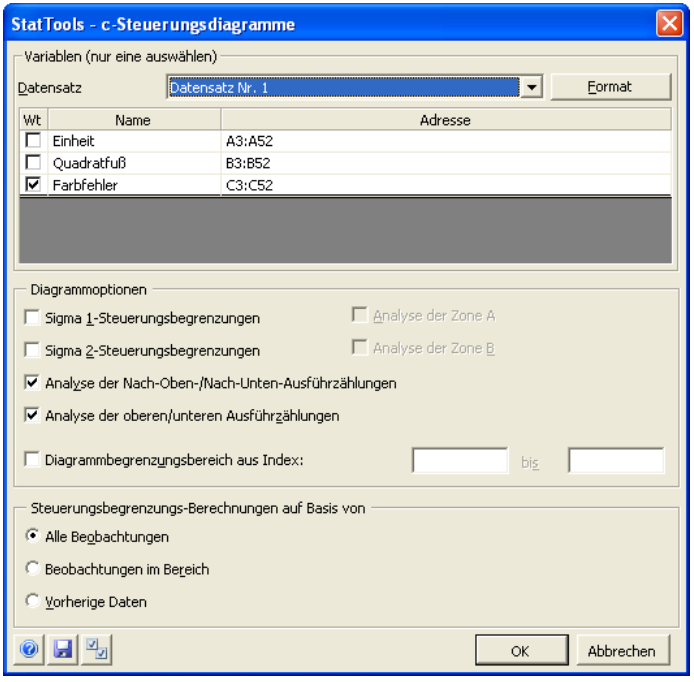
Erstellt *c*-Diagramme für Zeitserien-Variablen

c-Diagramme werden dazu verwendet, die Anzahl der Defekte bei Elementen gleicher Größe grafisch darzustellen. Angenommen, die Autotüren werden in Einzelfertigungen von jeweils 50 Türen hergestellt. Für jede Einzelfertigung würde dann die Anzahl der Defekte (z.B. Farbfehler oder scharfe Kanten) gezählt werden. Diese Zählungswerte werden dann in einem Diagramm grafisch dargestellt. Wie immer, soll durch diesen Vorgang festgestellt werden, ob der Prozess richtig erfasst oder gesteuert wird.

Der Datensatz für ein *c*-Diagramm muss eine Variable enthalten, aus der der Zählungswert für die Defekte in den einzelnen Elementen hervorgeht. Bei jeder Beobachtung wird davon ausgegangen, dass Elementgröße immer gleich ist. Wenn es sich bei einem „Element" beispielsweise um eine Kleinserienfertigung von Autotüren handelt, würden wir annehmen, dass jede dieser Serienfertigungen aus der gleichen Anzahl an Autotüren besteht.

***Dialogfeld* c-Steuerdiagramme**

Zum Definieren dieses Diagramms wird das Dialogfeld **c-Steuerdiagramme** verwendet:



Für diese Analyse werden eine oder mehrere Variablen ausgewählt. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

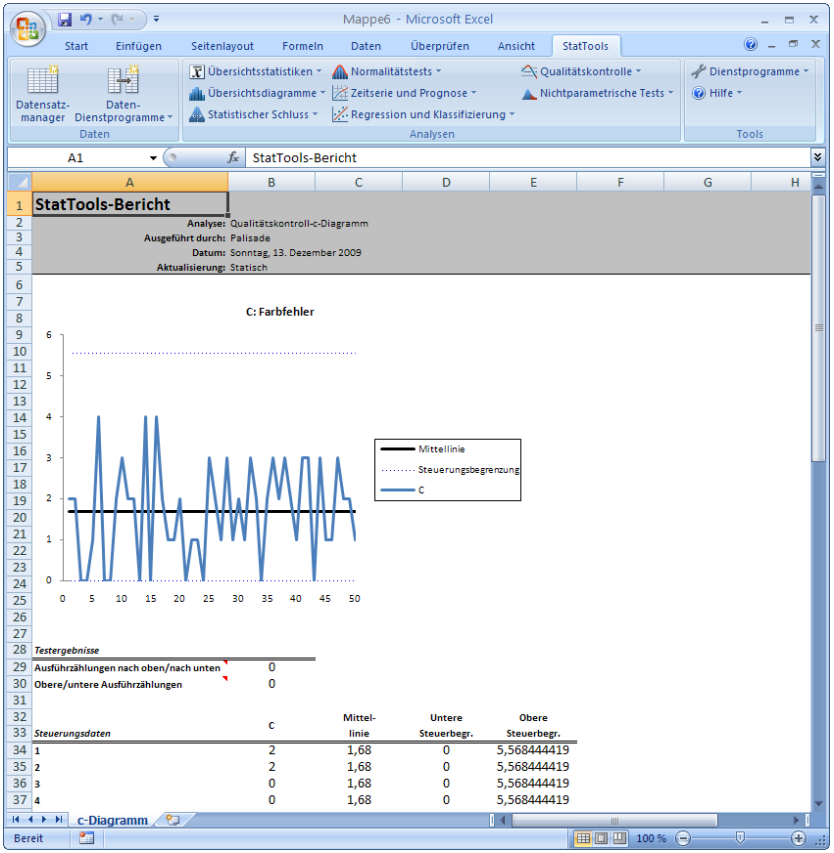
Unter **Diagrammoptionen** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Sigma 1-** und **Sigma 2-Steuergrenzen** – fügt bei Sigma 1 oder Sigma 2 von der Mittellinie gewisse Steuergrenzlinien hinzu Diese zusätzlichen Linien ermöglichen Ihnen, nach anderem nicht erfassten Verhalten (den so genannten „Zonenregeln“) zu suchen.
- **Analyse der Zone A** und **Zone B** – Anzahl der Punkte außerhalb der Zone A (Sigma 2) und Zone B (Sigma 1).
- **Analyse der Nach-Oben-/Nach-Unten-Ausführzählungen** und **Analyse der oberen/unteren Ausführzählungen** – sequenzielle Längenbewegungen nach oben oder unten im Werte von 8 oder mehr
- **Diagrammbegrenzungsbereich aus Index** – begrenzt die Punkte für eine Variable im Diagramm auf einen Datenpunktebereich (d.h., auf den Bereich vom Start- bis zum Endindex)

Steuerbegrenzungs-Berechnungen auf Basis von – bestimmt in diesem Dialogfeld die Daten, die als Basis für die Steuerbegrenzungs-Berechnungen dienen. Folgende Optionen sind verfügbar:

- **Alle Beobachtungen** – verwendet alle verfügbaren Daten in den Steuerbegrenzungs-Berechnungen
- **Beobachtungen im Bereich** – verwendet in den Steuerbegrenzungs-Berechnungen die Daten zwischen **Startindex** und **Stoppindex**
- **Vorherige Daten** – erstellt Steuerbegrenzungen aus den vorhergehend beobachteten Daten. Zu diesem Zweck braucht nur der *c*-Durchschnitt eingegeben werden, der für die vorherigen Daten berechnet wurde.

c-Diagramm (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten können nicht gehandhabt werden.
- **Mit Daten verknüpfen** – Diagramme sind nicht mit den Daten verknüpft.

Befehl *u-Diagramm*

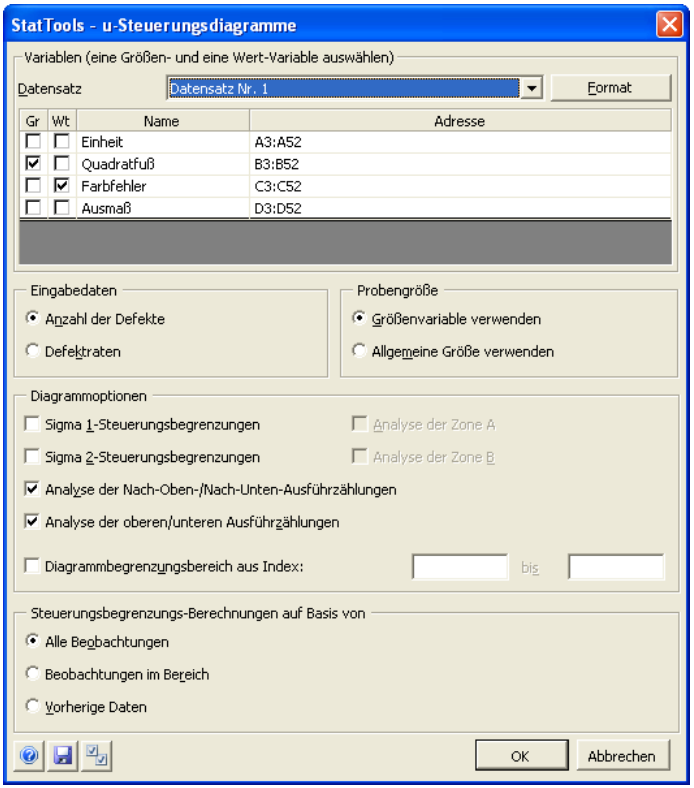
Erstellt *u*-Diagramme für Zeitserien-Variablen

u-Diagramme sind so ähnlich wie *c*-Diagramme, stellen aber die Defektrate dar. Nehmen wir einmal an, dass bei den Autotüren die Kleinserienfertigungen nicht immer die gleiche Anzahl an Türen enthalten. In diesem Fall würden wir mithilfe eines *u*-Diagramms die Defektrate pro Autotür darstellen. Dabei handelt es sich um die Anzahl der Defekte in einer Kleinserienfertigung dividiert durch die Anzahl der Türen pro Fertigung. Wie immer, soll durch diesen Vorgang festgestellt werden, ob der Prozess richtig erfasst oder gesteuert wird.

Für dieses Verfahren ist ein Datensatz erforderlich, der mindestens eine Variable enthält, aus der die *Anzahl* der Defekte für die einzelnen Beobachtungen hervorgeht oder die *Defektrate* pro Beobachtung. Wahlweise kann auch eine Variable vorhanden sein, aus der die Elementgrößen hervorgehen. Wenn solche Variable nicht verwendet wird, müssen Sie eine Elementgröße eingeben, von der dann angenommen wird, dass sie für alle Beobachtungen gleich ist. Wenn jedoch eine Größenvariable verfügbar ist, kann mit unterschiedlichen Elementgrößen gearbeitet werden.

***Dialogfeld* u-Steuerdiagramme**

Zum Definieren dieses Diagramms wird das Dialogfeld **u-Steuerdiagramme** verwendet:



Hier müssen zum Analysieren eine Wertvariable und (wahlweise) eine Variable für die Probengröße ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um entstapelte Daten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Unter **Eingabedaten** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Anzahl der Defekte** – legt fest, dass durch die Wertvariable die effektive Anzahl der nicht normgerechten Elemente in der gesamten Werteprobe angegeben wird
- **Defektraten** – legt fest, dass durch die Wertvariable der Anteil der nicht normgerechten Elemente in der Werteprobe angegeben wird

Unter **Probengröße** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Größenvariable verwenden** – legt fest, dass eine Größenvariable verwendet wird, um die Gesamtgröße der einzelnen Werteproben anzugeben
- **Allgemeine Größe verwenden** – legt fest, dass keine Größenvariable verwendet wird, da die einzelnen Werteproben der eingegebenen Größe entsprechen

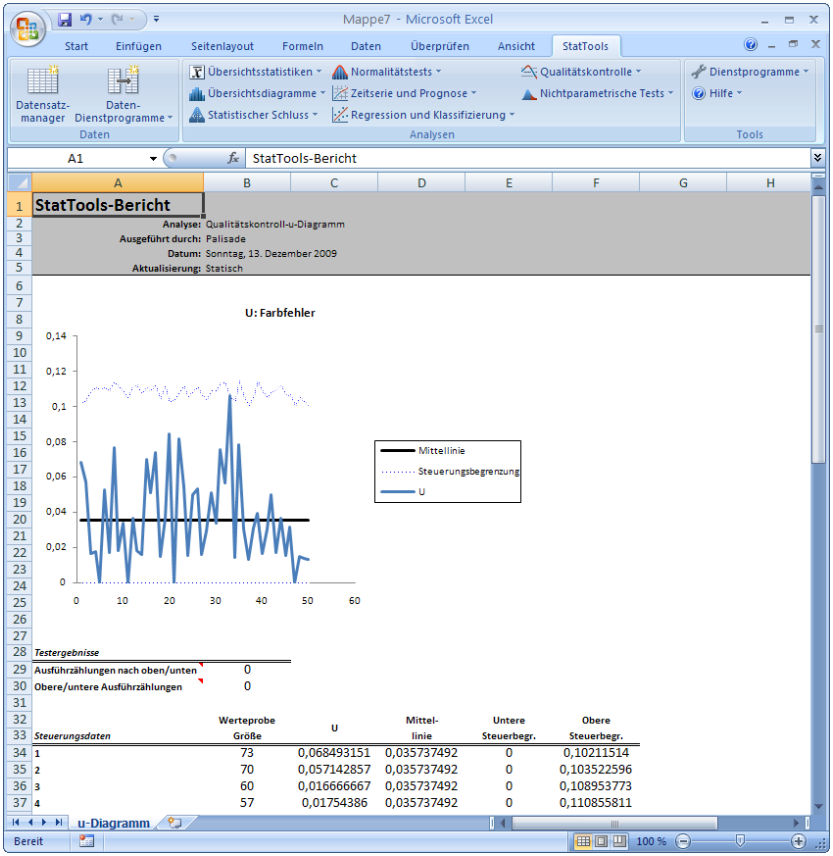
Unter **Diagrammoptionen** sind in diesem Dialogfeld folgende Optionen verfügbar:

- **Sigma 1-** und **Sigma 2-Steuergrenzen** – fügt bei Sigma 1 oder Sigma 2 von der Mittellinie gewisse Steuergrenzlinien hinzu Diese zusätzlichen Linien ermöglichen Ihnen, nach anderem nicht erfassten Verhalten (den so genannten „Zonenregeln“) zu suchen.
- **Analyse der Zone A** und **Zone B** – Anzahl der Punkte außerhalb der Zone A (Sigma 2) und Zone B (Sigma 1).
- **Analyse der Nach-Oben-/Nach-Unten-Ausführzählungen** und **Analyse der oberen/unteren Ausführzählungen** – sequenzielle Längenbewegungen nach oben oder unten im Werte von 8 oder mehr
- **Diagrammbegrenzungsbereich aus Index** – begrenzt die Punkte für eine Variable im Diagramm auf einen Datenpunktebereich (d.h., auf den Bereich vom Start- bis zum Endindex)

Steuerbegrenzungs-Berechnungen auf Basis von – bestimmt in diesem Dialogfeld die Daten, die als Basis für die Steuerbegrenzungs-Berechnungen dienen. Folgende Optionen sind verfügbar:

- **Alle Beobachtungen** – verwendet alle verfügbaren Daten in den Steuerbegrenzungs-Berechnungen
- **Beobachtungen im Bereich** – verwendet in den Steuerbegrenzungs-Berechnungen die Daten zwischen **Startindex** und **Stoppindex**
- **Vorherige Daten** – erstellt Steuerbegrenzungen aus den vorhergehend beobachteten Daten. Zu diesem Zweck brauchen nur **Größe der Probenuntermenge** und **u-Durchschnitt** der vorhergehend berechneten Daten eingegeben werden.

u-Diagramm (Beispiel)



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten können nicht gehandhabt werden.
- **Mit Daten verknüpfen** – Diagramme sind nicht mit den Daten verknüpft.

Menü *Nichtparametrische Tests*

„Nichtparametrische“ Tests sind statistische Vorgänge, die auf Datenproben angewendet werden, um Hypothesen über grundlegende Wahrscheinlichkeitsverteilungen zu untersuchen. „Parametrische“ Hypothesetests sind gebräuchlicher und bekannter, aber die nichtparametrische Alternative bietet Vorteile, die diese in vielen Situationen geeigneter erscheinen lässt.

Bei parametrischen Hypothesetests wird meistens von der Annahme ausgegangen, dass es sich um eine Normalverteilung handelt, und werden dann die zugehörigen Verteilungsparameter geschätzt, bei denen es sich typischerweise um den Mittelwert und die Standardabweichung handelt. Aber bei vielen Anwendungen ist die Verteilung nicht unbedingt normal. Die Anzahl der Anrufe pro Stunde in einem Kundenberatungszentrum und die Wartezeit an der Kasse im Supermarkt sind beispielsweise Dinge, die nicht der Normalverteilung folgen. Nichtparametrische Tests sind in solchen Fällen geeigneter, da sie keine bestimmte Verteilung erforderlich machen. Einige von diesen Tests gehen jedoch von allgemeinen Annahmen hinsichtlich der Verteilungsform aus. In diesem Paket wird im Wilcoxon-Vorzeichen-Rangtest davon ausgegangen, dass die Verteilung symmetrisch ist. In den beiden anderen in diesem Paket befindlichen Tests (dem Vorzeichentest und dem Mann-Whitney-Test) wird nicht von irgendeiner bestimmten Verteilungsform ausgegangen.

Bei kleinen Werteprobengrößen ist die nichtparametrische Alternative oft geeigneter. Wenn die Werteprobe dagegen groß ist, kann ein Normalitätstest angewendet werden. Falls die Annahme, dass die Verteilung normal ist, tatsächlich zutrifft, ist ein parametrischer Test durchaus angebracht. Bei kleinen Probengrößen kann durch Normalitätstests jedoch kaum zwischen normalen und anderen Verteilungen unterschieden werden. In solchen Fällen können nichtparametrische Tests aus dem Dilemma helfen.

Ordnungsdaten

Für bestimmte Datentypen können keine parametrischen, aber einige nichtparametrische Tests verwendet werden. Das ist z.B. bei **Ordnungsdaten** der Fall, bei denen Beobachtungen in Form von Nummern beschrieben werden, durch die auf eine Rangfolge hingewiesen wird. Der Unterschied zwischen zwei solcher Nummern ist jedoch nicht sehr aussagefähig. Bildungsabschlüsse können beispielsweise zwischen 0 (nur Grundschule), 1 (Hochschule ohne Abschluss), 2 (Hochschule mit Abschlusszeugnis), 3 (College ohne Abschluss), 4 (College mit Abschlusszeugnis) und 5 (Diplomausbildung) eingestuft werden. Wenn diese Einstufung verwendet wird, bedeutet das nicht, dass die Differenz im Bildungsabschluss zwischen „Hochschule mit Abschlusszeugnis“ und „Grundschule“ dem Unterschied zwischen den Kategorien „Diplomausbildung“ und „College ohne Abschluss“ entspricht, obwohl die Differenz in beiden Fällen gleich 2 ist. Dieses Paket enthält Tests, die auf solche Daten angewendet werden können – z.B. den Vorzeichentest (Analyse mit einer Werteprobe) und den Mann-Whitney-Test.

Verwendungs-übersicht

Zusammenfassend ist zu sagen, dass nichtparametrische Tests in folgenden Situationen angewendet werden können, in denen parametrische Tests nicht geeignet sind:

- bei geringen Informationen über die zugrunde liegende Wahrscheinlichkeitsverteilung
- wenn die Probengröße zu klein ist, um die Normalität zuverlässig prüfen zu können
- wenn es sich bei den Daten um Ordnungsdaten handelt

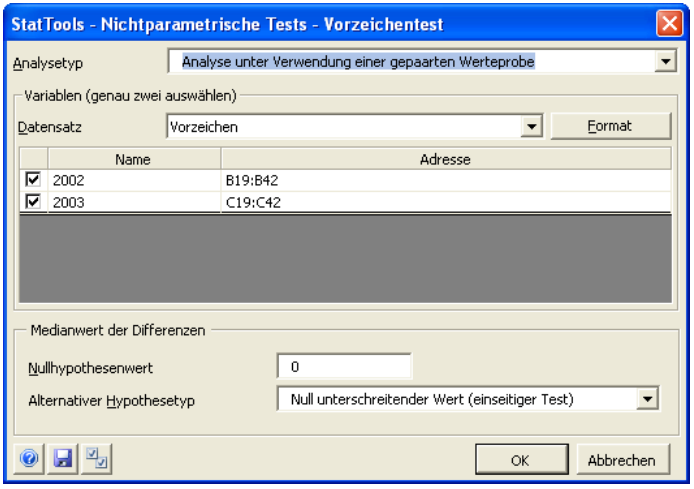
Befehl *Vorzeichentest*

Führt den Vorzeichentest für Variablen aus

Über diesen Befehl können Hypothesetests für den Medianwert einer einzelnen Variable ausgeführt werden (Analyse mit einer Werteprobe) oder auch für den Medianwert der Differenzen bei gepaarten Variablen (Analyse mit gepaarter Werteprobe). Bei diesem Test wird nicht von einer bestimmten Verteilungsform ausgegangen (insbesondere ist keine Normalverteilung erforderlich). Wie in de Übersicht beschrieben, kann die Analyse mit einer Werteprobe für **Ordnungsdaten** verwendet werden.

***Dialogfeld* Vorzeichentest**

Zum Definieren dieser Analyse wird das Dialogfeld **Vorzeichentest** verwendet:

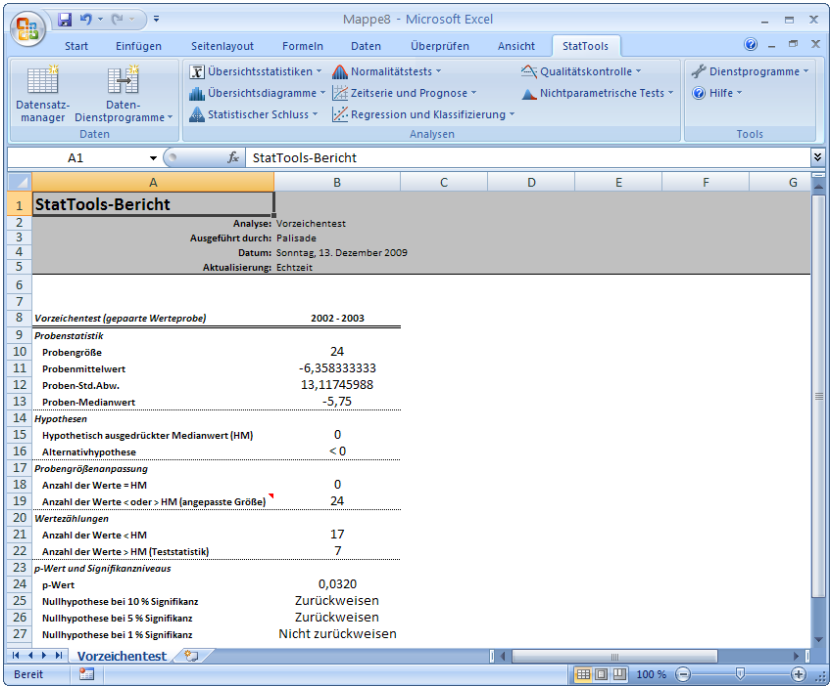


Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Die Analyse unter Verwendung nur einer Werteprobe erfordert eine oder mehrere Variablen, während für eine Analyse mit gepaarter Werteprobe zwei Variablen benötigt werden. Bei einer Analyse mit nur einer Werteprobe können die ausgewählten Variablen aus gestapelten oder entstapelten Daten bestehen. Bei der Analyse mit gepaarter Werteprobe müssen die Daten dagegen entstapelt sein. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Dabei sind folgende Optionen verfügbar:
  - **Analyse mit einer Werteprobe –** führt Hypothesetests für eine einzelne numerische Variable aus.
  - **Analyse mit gepaarter Werteprobe –** Diese Analyse sollte für natürlich gepaarte Variablen verwendet werden. Mit anderen Worten, diese Analyse entspricht einer Analyse mit einer Werteprobe in Bezug auf die Differenzen zwischen Variablenpaaren.
- Optionen unter **Medianwert der Differenzen**
  - **Nullhypothesenwert –** d.h., der Wert des Populationsparameters unter der Nullhypothese
  - **Alternativer Hypothesetyp –** d.h., die Alternative zum Nullhypothesenwert, der während der Analyse ausgewertet wird. Die alternative Hypothese kann entweder **einseitig** (d.h. größer oder kleiner als die Nullhypothese) oder **zweiseitig** (d.h., ungleich der Nullhypothese) sein.

Bericht über Vorzeichentest



Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten** – Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen** – Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

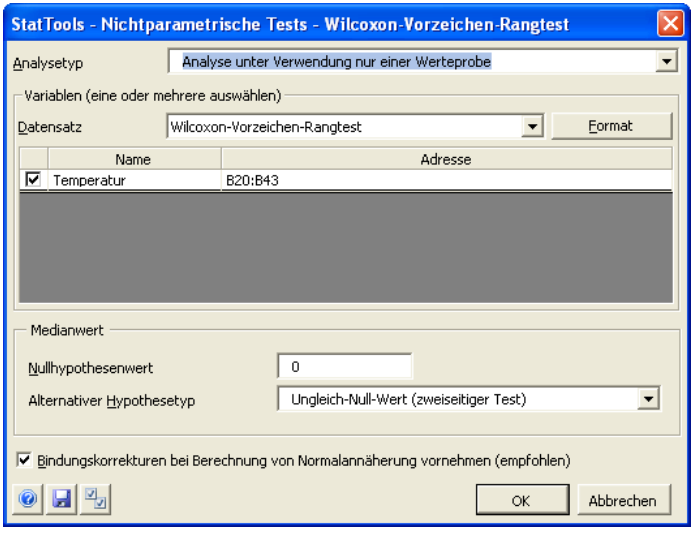
Befehl *Wilcoxon-Vorzeichen-Rangtest*

Führt den Wilcoxon-Vorzeichen-Rangtest an Variablen aus

Über diesen Befehl können Hypothesetests für den Medianwert einer einzelnen Variable ausgeführt werden (Analyse mit einer Werteprobe) oder auch für den Medianwert der Differenzen bei gepaarten Variablen (Analyse mit gepaarter Werteprobe). Bei diesem Test wird davon ausgegangen, dass die Wahrscheinlichkeitsverteilung symmetrisch (aber nicht unbedingt normal) ist.

Zum Definieren dieser Analyse wird das Dialogfeld **Wilcoxon-Vorzeichen-Rangtest** verwendet:

***Dialogfeld* Wilcoxon-Vorzeichen-Rangtest**

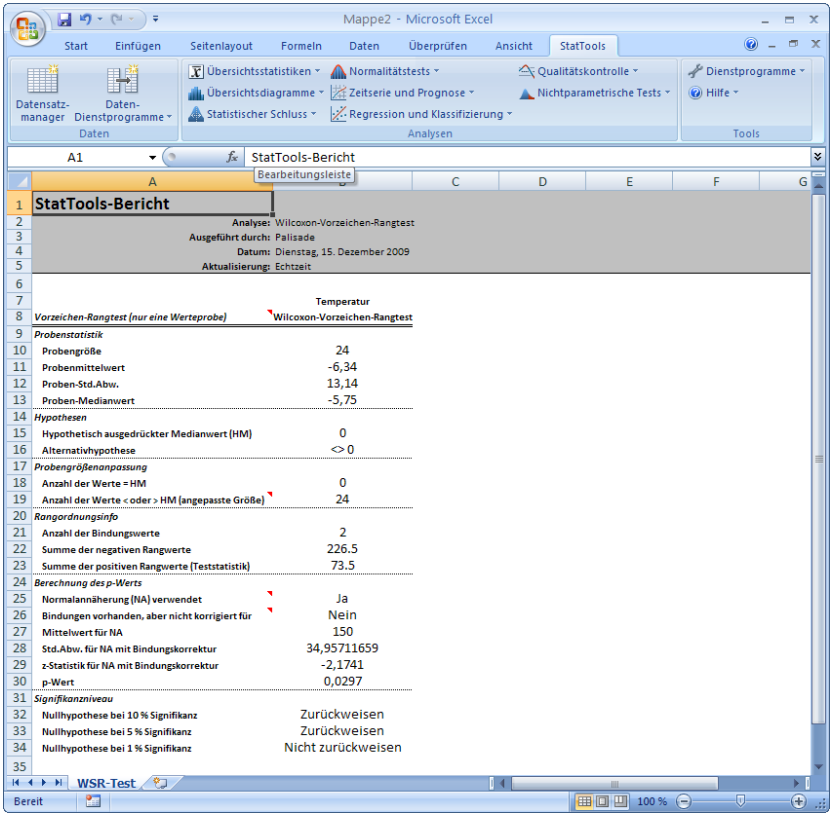


Die Anzahl der ausgewählten Variablen hängt vom verwendeten Analysetyp ab. Die Analyse unter Verwendung nur einer Werteprobe erfordert eine oder mehrere Variablen, während für eine Analyse mit gepaarter Werteprobe zwei Variablen benötigt werden. Bei einer Analyse mit nur einer Werteprobe können die ausgewählten Variablen aus gestapelten oder entstapelten Daten bestehen. Bei der Analyse mit gepaarter Werteprobe müssen die Daten dagegen entstapelt sein. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

In diesem Dialogfeld sind folgende Optionen verfügbar:

- **Analysetyp –** wählt die Art der auszuführenden Analyse aus. Dabei sind folgende Optionen verfügbar:
  - **Analyse mit einer Werteprobe –** führt Hypothesetests für eine einzelne numerische Variable aus.
  - **Analyse mit gepaarter Werteprobe –** Dieser Analysetyp sollte für natürlich gepaarte Variablen verwendet werden. Mit anderen Worten, diese Analyse entspricht einer Analyse mit einer Werteprobe in Bezug auf die Differenzen zwischen Variablenpaaren.
- Optionen unter **Medianwert**
  - **Nullhypothesenwert –** d.h., der Wert des Populationsparameters unter der Nullhypothese
  - **Alternativer Hypothesetyp –** d.h., die Alternative zum Nullhypothesenwert, der während der Analyse ausgewertet wird. Die alternative Hypothese kann entweder **einseitig** (d.h. größer oder kleiner als die Nullhypothese) oder **zweiseitig** (d.h., ungleich der Nullhypothese) sein.
- **Bindungskorrektur –** Dies ist eine empfohlene Auswahl, durch die im Test die gebundenen Rangfolgen korrigiert werden, aber nur bei Verwendung der normalen Annäherungswerte. Bei dieser Korrektur handelt es sich um das Zählen der Elemente in Gruppen von gebundenen Rangfolgen und das entsprechende Reduzieren der Varianz. Durch diese Bindungskorrektur wird stets der Wert der *z*-Statistik erhöht, sofern gebundene Rangfolgen vorhanden sind. (Hinweis: Die Bindungskorrektur verursacht keine Änderung in der Varianz, wenn keine Bindungen vorhanden sind.)

Bericht über Wilcoxon-Vorzeichen-Rangtest



Hinweis: Wenn die Probengröße den Wert 15 überschreitet, wird in diesem Bericht der p-Wert mithilfe der Normalannäherung berechnet.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

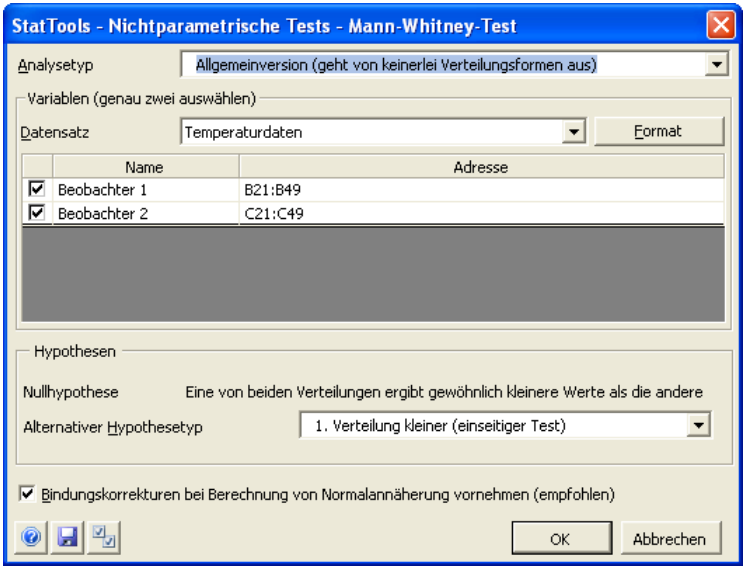
Befehl *Mann-Whitney-Test*

Führt den Mann-Whitney-Test an Variablen aus

Über diesen Befehl kann ein Hypothesetest an zwei Werteproben ausgeführt werden. In einer Version dieses Tests (der so genannten *Medianversion*) besagt die Hypothese, dass die Medianwerte von zwei Populationen identisch sind. In dieser Version wird davon ausgegangen, dass die Wahrscheinlichkeitsverteilungen ein und dieselbe Form haben. In der anderen Version (der so genannten *Allgemeinversion*) wird nicht von dieser Annahme ausgegangen und die Hypothese bestreitet, dass eine der Verteilungen gewöhnlich kleinere Werte als die andere ergibt. Mit anderen Worten, die Hypothese behauptet, dass P[X1>X2] = P[X2>X1] ist, wobei P[X1>X2] die Wahrscheinlichkeit darstellt, dass eine Beobachtung aus Population 1 größer als eine Beobachtung aus Population 2 ist. Hinweis: Der Mann-Whitney-Test kann zum Zurückweisen der Hypothese verwendet werden, dass zwei Werteproben aus derselben Wahrscheinlichkeitsverteilung stammen. Der Mann-Whitney-Test wird oft auch als Wilcoxon-Rangsummentest bezeichnet.

Dialogfeld Mann-Whitney-Test

Zum Definieren dieser Analyse wird das Dialogfeld **Mann-Whitney-Test** verwendet:



Für diese Analyse sind zwei Variablen erforderlich. Diese Variablen können gestapelt oder entstapelt sein und auch aus verschiedenen Datensätzen stammen.

Im Dialogfeld **Mann-Whitney-Test** sind folgende Optionen verfügbar:

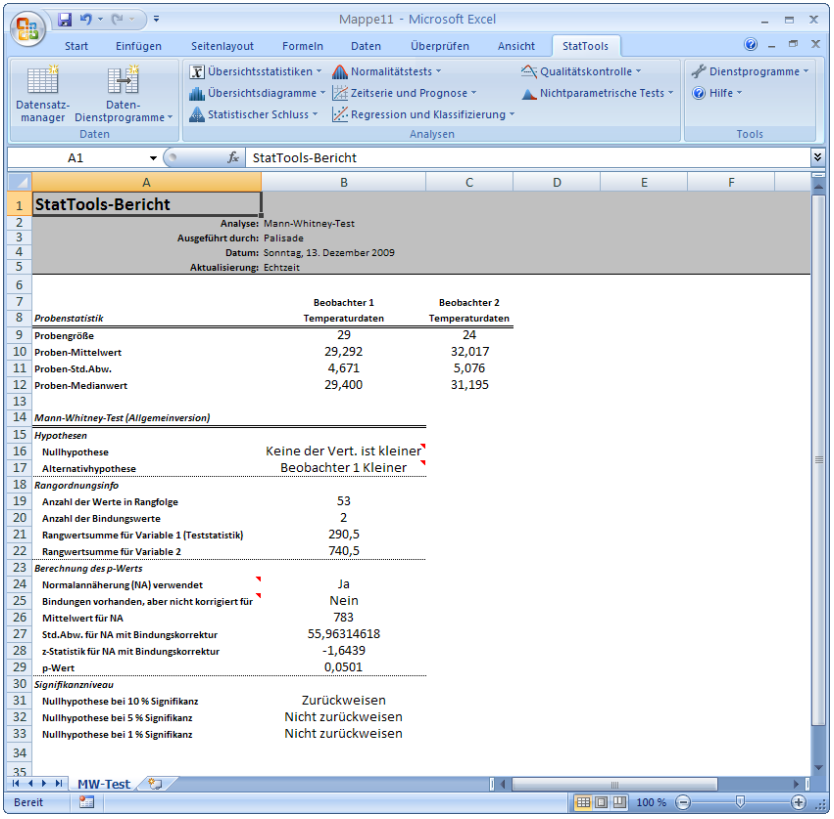
- **Analysetyp** – wählt die Formulierung der Nullhypothese und die Alternativen aus. Folgende Optionen können gewählt werden:
  - **Allgemeinversion** – führt einen Hypothesetest aus, um festzustellen, ob eine Wahrscheinlichkeitsverteilung dazu neigt, kleinere Werte zu erbringen als die andere.
    - **Hypothesen**:
      - **Nullhypothese** – bestreitet, dass eine der Wahrscheinlichkeitsverteilungen dazu neigt, kleinere Werte zu erbringen als die andere Verteilung. Mit anderen Worten, diese Hypothese erklärt, dass P[X1>X2] = P[X2>X1] ist, wobei P[X1>X2] die Wahrscheinlichkeit darstellt, dass eine Beobachtung aus Population 1 größer als eine Beobachtung aus Population 2 ist. P[X2>X1] ist analog auszulegen. Bei kontinuierlichen Verteilungen entspricht dieses der Aussage, dass beide Wahrscheinlichkeiten der Gleichung 0,5 (P[X1>X2] = P[X2>X1] = 0,5 entsprechen.
      - **Alternativer Hypothesetyp** – Die Alternativhypothese kann entweder **einseitig** (d.h., eine Wahrscheinlichkeit ist größer oder kleiner als die andere) oder **zweiseitig** (d.h., die beiden Wahrscheinlichkeiten sind ungleich) sein.

- **Medianversion –** führt einen Hypothesetest aus, um festzustellen, ob der Medianwert der einen Population dem Medianwert der anderen Population gleicht oder unterschiedlich ist. Dabei wird davon ausgegangen, dass beide Verteilungen die gleiche Form haben.
    - **Hypothesen**:
        - **Nullhypothese –** besagt, dass die beiden Medianwerte gleich sind.
        - **Alternativer Hypothesetyp –** Die Alternativhypothese kann entweder **einseitig** (d.h., der Medianwert der einen Population ist größer oder kleiner als der Medianwert der anderen Population) oder **zweiseitig** (d.h., die beiden Medianwerte sind ungleich) sein.

Hinweis: Die in der Allgemeinversion und Medianversion des Tests ausgeführten Berechnungen sind die gleichen. Die beiden Versionen unterscheiden sich nur hinsichtlich der Annahmen in Bezug auf einheitliche Verteilungsformen und Nullhypothese. Diese beiden Versionen sind dazu da, um klarzustellen, dass der Mann-Whitney-Test selbst dann angewendet werden kann, wenn nicht anzunehmen ist, dass die Verteilungen ungefähr gleicher Form sind, solange eine geeignete Nullhypothese in Betracht gezogen wird. Wenn Sie den Test bei zwei Verteilungen mit deutlich unterschiedlichen Formen ausführen und die Nullhypothese durch den Test zurückgewiesen wird, könnte der Grund dafür in unterschiedlichen Medianwerten liegen. Es könnten aber auch verschiedene Varianzen vorhanden sein oder noch andere Ursachen vorliegen.

- **Bindungskorrektur –** Dies ist eine empfohlene Auswahl, durch die im Test die verbundenen Rangfolgen korrigiert werden, aber nur bei Verwendung der normalen Annäherungswerte. Bei dieser Korrektur handelt es sich um das Zählen der Elemente in Gruppen von gebundenen Rangfolgen und das entsprechende Reduzieren der Varianz. Durch diese Bindungskorrektur wird stets der Wert der *z*-Statistik erhöht, sofern gebundene Rangfolgen vorhanden sind. (Hinweis: Die Bindungskorrektur verursacht keine Änderung in der Varianz, wenn keine Bindungen vorhanden sind.)

Bericht über Mann-Whitney-Test



Wenn die Größe einer der beiden Proben den Wert 10 überschreitet, wird im vorstehenden Bericht der *p-Wert* mittels Normalannäherung berechnet (aber nicht, wenn eine der Größen 11 oder 12 entspricht, während die andere bei 3 oder 4 liegt).

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Alle Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variable ändern, werden die Ausgaben automatisch entsprechend aktualisiert.

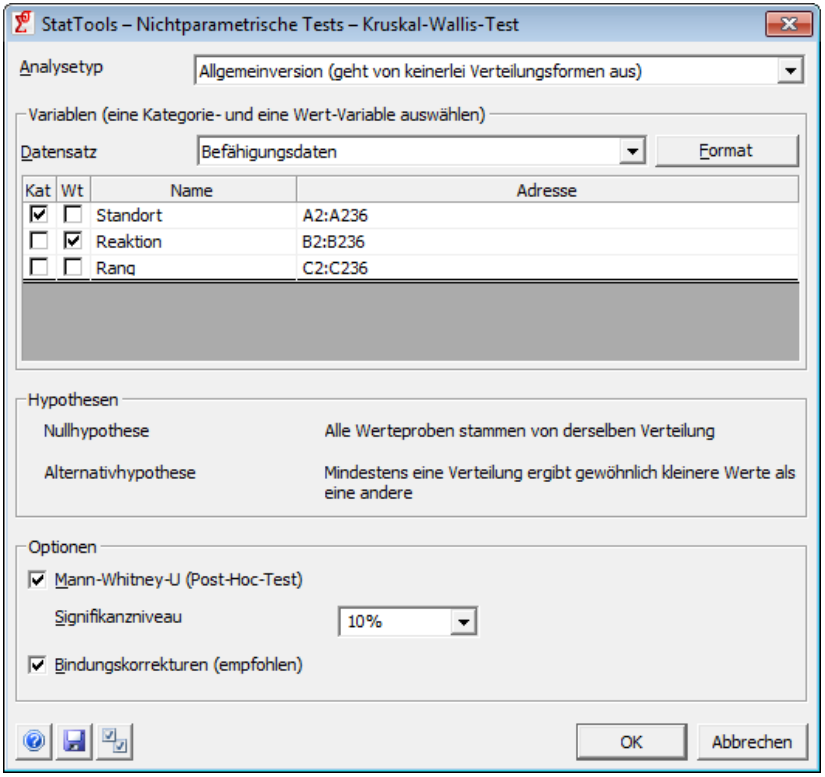
Befehl Kruskal-Wallis-Test

Führt den Kruskal-Wallis-Test an Variablen aus

Über den Befehl Kruskal-Wallis-Test kann ein Hypothesetest an drei oder mehr Werteproben ausgeführt werden. Er erweitert den **Mann–Whitney**-Test, der für Variablenpaare angewandt wird. Bei der Allgemeinversion des Tests gibt die Hypothese an, dass alle Werteproben aus der gleichen Verteilung stammen. In diesem Fall werden kontinuierliche Verteilungen vorausgesetzt. Andererseits könnte es möglich sein, dass alle Populationsverteilungen die gleiche Form haben (normal oder nicht); die Hypothese lautet dann, dass die Medianwerte der Populationen identisch sind.

Zum Definieren dieser Analyse wird das Dialogfeld Kruskal-Wallis verwendet:

Dialogfeld Kruskal-Wallis-Test



Es können drei oder mehr Variablen ausgewählt werden. Der Datensatz kann gestapelt oder entstapelt sein. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Im Dialogfeld Kruskal-Wallis-Test sind folgende Optionen verfügbar:

- **Analysetyp –** Wählt die Art der auszuführenden Analyse aus. Hierfür sind folgende Optionen verfügbar:
  - **Allgemeinversion –** Sie setzt voraus, dass die Wertevariable in jeder Population eine kontinuierliche Verteilung hat und die Nullhypothese besagt, dass alle Werteproben aus derselben Verteilung stammen.
  - **Medianversion –** Sie setzt voraus, dass alle Populationsverteilungen die gleiche Form haben und die Nullhypothese besagt, dass die Medianwerte der Populationen identisch sind.
- **Mann-Whitney-U-Test –** Der Kruskal-Wallis-Test gibt – so wie ein einseitiger ANOVA-Test – an, ob es beachtliche Unterschiede zwischen Kategorien gibt. Er sagt jedoch nichts darüber aus, welche Kategorien unterschiedlich sind. StatTools bietet daher eine Option zur Verwendung des Mann-Whitney-U-Tests mit der Bonferroni-Korrektur. Die Bonferroni-Korrektur ist eine Anpassung der p-Werte, wenn an denselben Daten mehrere statistische Tests gleichzeitig ausgeführt werden.
- **Bindungskorrekturen** – Da die Varianz der Werteprobenverteilung der KW-Teststatistik durch Ranggleichheiten beeinflusst wird, können Ranggleichheiten in der Berechnung dieser Teststatistiken korrigiert werden.

Kruskal-Wallis-Testbericht

StatTools-Bericht

Analyse: Kruskal-Wallis-Test
Ausgeführt durch: Palisade
Datum: Donnerstag, 18. Juni 2015
Aktualisierung: Echtzeit

Probenstatistik	Rang (Midwest) Befähigungsdaten	Rang (Northeast) Befähigungsdaten	Rang (South) Befähigungsdaten	Rang (Southwest) Befähigungsdaten	Rang (West) Befähigungsdaten
Probengröße	55	43	40	47	50
Probenmittelwert	119,15	156,88	108,90	73,87	132,06
Proben-Std.Abw.	74,77	57,73	72,47	52,25	53,34
Proben-Medianwert	135,00	170,00	99,00	64,00	129,50

Kruskal-Wallis-Test (Allgemeine Version)	
Hypothesen	
Nullhypothese	Gleiche Verteilung
Alternativhypothese	Verschiedene Verteilungen
Rangordnungsinfo	
Anzahl der Werte in Rangfolge	235
Bindungskorrektur aktiviert	WAHR
Anzahl der Bindungswerte	0
Teststatistik H	36,7407
Übersicht über Berechnung des p-Werts	
Berechnungsmethode	Chi-Quadrat-Annäherung
p-Wert	< 0.0001
Exakte Berechnung	
p-Wert (exakt)	#NV
Monte Carlo-Annäherung	
p-Wert (Monte Carlo)	#NV
Untergrenze des Vertrauensbereichs	#NV
Obergrenze des Vertrauensbereichs	#NV
Chi-Quadrat-Annäherung	
Freiheitsgrade	4
p-Wert (Chi-Quadrat)	< 0.0001
Signifikanzniveau	
Nullhypoth. bei 10 % Signifikanz	Zurückweisen
Nullhypoth. bei 5 % Signifikanz	Zurückweisen
Nullhypoth. bei 1 % Signifikanz	Zurückweisen

Bonferroni Mann-Whitney U-Test

Post-Hoc-Tests	Teststatistik	p-Wert	ypoth. bei 10% Signifikanz
Rang (Midwest) - Rang (Northeast)	2374,0000	0,0127	

In diesem Bericht wird der p-Wert mithilfe der Chi-Quadrat-Annäherung berechnet. StatTools hat drei verschiedene Methoden zur Berechnung des p–Werts:

- **Chi-Quadrat-Annäherung –** Der p-Wert wird mit der Annahme berechnet, dass die Werteprobenverteilung der KW-Teststatistik durch die Chi-Quadrat-Verteilung ungefähr angenähert wird. Die Annäherung verbessert sich, wenn sowohl die Anzahl der Variablen als auch die Anzahl der Beobachtungen in jeder Variable zunehmen.

- **Exakt –** Die Berechnung des p-Werts basiert auf der wahren Verteilung der KW-Teststatistik. Diese Berechnungen sind sehr umfangreich, da die Ränge der Werteprobendaten auf alle möglichen Weisen neu angeordnet werden müssen (der Wert der Teststatistik muss mit jedem möglichen Szenario berechnet werden). Dies wird daher nur für kleine Werteproben verwendet.
- **Monte Carlo –** Dies bietet eine Schätzung des exakten p-Werts ohne Beeinflussung und zeigt mit Hilfe der Monte Carlo Probenerhebungsmethode einen Vertrauensbereich an. Das wird verwendet, wenn die Datensätze für die exakte Methode zu groß sind oder wenn die für die Chi-Quadrat-Annäherungsmethode erforderlichen Annahmen nicht erfüllt werden.

  Es ist zu beachten, dass StatTools die exakte Methode verwendet, wenn die Anzahl der Variablen $k \leq 4$ und die Anzahl der Beobachtungen in den einzelnen Variablen $ni \leq 4$ beträgt. Die Monte Carlo Methode wird verwendet, wenn die exakte Methode nicht verwendet werden kann und eine der folgenden Bedingungen erfüllt wird: Es gibt mehr als drei Variablen ($k > 3$) wobei mindestens eine Variable eine kleine Anzahl von Beobachtungen enthält ($ni \leq 5$), oder es gibt exakt drei Variablen mit insgesamt $N \leq 30$ Beobachtungen in allen drei Variablen. Die Monte Carlo Methode führt 10000 Iterationen (d. h., es werden 10000 Werteproben ausgewählt) aus und kann Simulationen mit einer Gesamtanzahl von Beobachtungen in allen Variablen von $N \leq 25000$ unterstützen. Wenn die exakte oder die Monte Carlo Methode nicht verwendet werden kann, wird die Chi-Quadrat-Annäherung verwendet.

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zellen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Die Live-Berichte werden durch Formeln berechnet, die mit den Daten verknüpft sind. Wenn sich die Werte der ausgewählten Variablen ändern, ändern sich die angezeigten Werte automatisch; die Methode der p-Wert-Berechnung ändert sich bei Bedarf auch automatisch.

Menü *multivariate Analyse*

Befehl Hauptkomponenten-Analyse

Führt an ausgewählten Variablen die Hauptkomponenten-Analyse aus

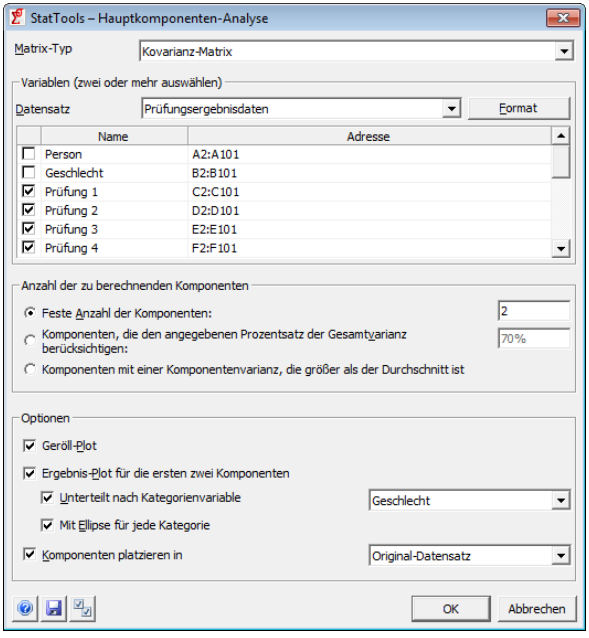
Die Hauptkomponenten-Analyse ist ein Verfahren zur Verringerung der Dimensionalität eines multivariaten Datensatzes. Sie transformiert einen Satz korrelierter Variablen in eine kleinere Anzahl nicht korrelierter Variablen.

Diese Analyse wird oft als Eingabe für andere Analysen verwendet:

- Bei der linearen Regression wird sie verwendet, wenn es Multikollinearitätsprobleme gibt oder wenn die Anzahl der unabhängigen Variablen im Vergleich zur Anzahl der Beobachtungen relativ groß ist.
- Sie ist zum Erstellen eines Punktdiagramms der ersten zwei Hauptkomponenten nützlich, um nach Ausreißern und dem Vorhandensein von Clustern zu prüfen.

Zum Definieren dieser Analyse wird das Dialogfeld Hauptkomponenten-Analyse verwendet:

Dialogfeld Hauptkompo-nenten-Analyse



Es können zwei Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um einen entstapelten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Im Dialogfeld Hauptkomponenten-Analyse stehen folgende Optionen zur Verfügung:

- **Matrix-Typ –** Wählt die Matrix aus, aus der die Hauptkomponenten extrahiert werden. Es sind folgende Optionen verfügbar:
  - **Kovarianz-Matrix –** Berechnet die Hauptkomponenten aus einer Kovarianz-Matrix. Diese Option sollte verwendet werden, wenn die Maßeinheiten für die verschiedenen Variablen verhältnismäßig sind (beispielsweise sind Zentimeter und Zoll verhältnismäßig, Zentimeter und Grad Celsius sind nicht verhältnismäßig.)
  - **Korrelations-Matrix –** Berechnet die Hauptkomponenten aus einer Korrelations-Matrix. Verwenden Sie diese Option, wenn die Maßeinheiten für verschiedene Variablen nicht verhältnismäßig sind.

Im Abschnitt **Anzahl der zu berechnenden Komponenten** stehen folgende Optionen zur Verfügung:

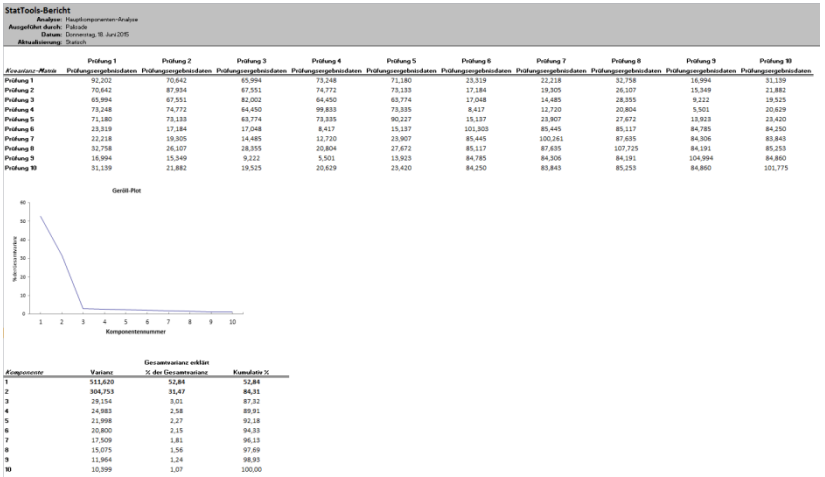
- **Feste Anzahl der Komponenten –** Legt die Anzahl der zu berechnenden Hauptkomponenten fest. Verwenden Sie diese Option, wenn Sie die Anzahl der Komponenten, die berechnet werden soll, bereits kennen.
- **Komponenten, die den angegebenen Prozentsatz der Gesamtvarianz berücksichtigen –** Behält eine ausreichende Anzahl von Komponenten bei, um einen angegebenen Prozentsatz der Gesamtvarianz zu berücksichtigen.
- **Komponenten mit einer Komponentenvarianz, die größer als der Durchschnitt ist –** Behält eine Komponente bei, wenn deren Varianz größer als der Durchschnitt der Varianzen aller Komponenten ist.

Es sind auch noch mehrere zusätzliche Optionen verfügbar:

- **Geröll-Plot –** Stellt den Prozentsatz der Gesamtvarianz grafisch dar, erläutert durch die einzelnen Komponenten.
- **Ergebnis-Plot für die ersten zwei Komponenten** – Stellt die erste Hauptkomponente im Vergleich zur zweiten Hauptkomponente grafisch dar.

- **Unterteilt nach Kategorienvariable –** Mit dieser Option wird ein Punktdiagramm für die ersten zwei Hauptkomponenten erstellt, nach einer Kategorien-Variable gefärbt.
- **Mit Ellipse für jede Kategorie –** Diese Option dient zum Erstellen einer Ellipse zu Cluster-Punkten der gleichen Kategorie.

Geröll-Plot



StatTools-Bericht

Analyse: Hauptkomponenten-Analyse
Ausgeführt durch: Palisade
Datum: Donnerstag, 18. Juni 2015
Aktualisierung: Statisch

Kovarianz-Matrix	Prüfung 1 Prüfungsergebnisdaten	Prüfung 2 Prüfungsergebnisdaten	Prüfung 3 Prüfungsergebnisdaten	Prüfung 4 Prüfungsergebnisdaten	Prüfung 5 Prüfungsergebnisdaten	Prüfung 6 Prüfungsergebnisdaten	Prüfung 7 Prüfungsergebnisdaten	Prüfung 8 Prüfungsergebnisdaten	Prüfung 9 Prüfungsergebnisdaten	Prüfung 10 Prüfungsergebnisdaten
Prüfung 1	92,202	70,642	65,994	73,248	71,180	23,319	22,218	32,758	16,994	31,139
Prüfung 2	70,642	87,934	67,551	74,772	73,133	17,184	19,305	26,107	15,349	21,882
Prüfung 3	65,994	67,551	82,002	64,450	63,774	17,048	14,485	28,355	9,222	19,525
Prüfung 4	73,248	74,772	64,450	99,833	73,335	8,417	12,720	20,804	5,501	20,629
Prüfung 5	71,180	73,133	63,774	73,335	90,227	15,137	23,907	27,672	13,923	23,420
Prüfung 6	23,319	17,184	17,048	8,417	15,137	101,303	85,445	85,117	84,785	84,250
Prüfung 7	22,218	19,305	14,485	12,720	23,907	85,445	100,261	87,635	84,306	83,843
Prüfung 8	32,758	26,107	28,355	20,804	27,672	85,117	87,635	107,725	84,191	85,253
Prüfung 9	16,994	15,349	9,222	5,501	13,923	84,785	84,306	84,191	104,994	84,860
Prüfung 10	31,139	21,882	19,525	20,629	23,420	84,250	83,843	85,253	84,860	101,775

Geröll-Plot

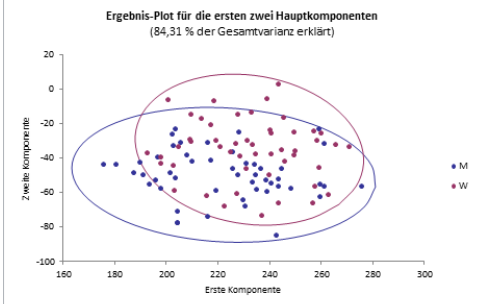
Gesamtvarianz erklärt

Komponente	Varianz	% der Gesamtvarianz	Kumulative %
1	511,620	52,84	52,84
2	304,753	31,47	84,31
3	29,154	3,01	87,32
4	24,983	2,58	89,91
5	21,998	2,27	92,18
6	20,800	2,15	94,33
7	17,509	1,81	96,13
8	15,075	1,56	97,69
9	11,964	1,24	98,93
10	10,399	1,07	100,00

Ergebnis-Plot

Koeffizienten-Matrix

Komponente	Prüfung 1	Prüfung 2	Prüfung 3	Prüfung 4	Prüfung 5	Prüfung 6	Prüfung 7	Prüfung 8	Prüfung 9	Prüfung 10
1	0,279	0,259	0,235	0,242	0,261	0,357	0,363	0,392	0,348	0,374
2	-0,344	-0,368	-0,339	-0,413	-0,362	0,274	0,259	0,212	0,302	0,227



Ergebnis Plot-Daten	Geschlecht	PC1	PC2
	M	207,9636	-38,2430
	M	236,4790	-45,7033
	M	259,1273	-23,1683

- **Komponenten platzieren in –** Legt fest, ob Variablen erzeugt werden, die die Hauptkomponenten darstellen. Wenn dies berücksichtigt wird, können Sie festlegen, wo diese Variablen platziert werden sollen: im ursprünglichen Datensatz oder in einem neuen. Wenn numerische Variablen aus verschiedenen Datensätzen ausgewählt werden, gibt es nur eine Wahlmöglichkeit: Komponenten in einen neuen Datensatz platzieren. Wenn die Option Komponenten platzieren in aktiviert ist, werden Spalten mit den Werten der Komponenten rechts des ausgewählten Datensatzes eingetragen.

Datensatz mit Hauptkompo-nenten

Geschlecht	Prüfung 1	Prüfung 2	Prüfung 3	Prüfung 4	Prüfung 5	Prüfung 6	Prüfung 7	Prüfung 8	Prüfung 9	Prüfung 10	PC1	PC2
W	75	75	73	68	74	80	88	90	83	84	249,2690	-25,0536
W	67	77	77	73	72	77	78	68	75	72	228,5904	-38,9530
M	73	67	70	66	63	73	65	61	71	62	207,9636	-38,2430
W	83	81	72	81	79	70	73	81	72	81	239,7710	-49,3879
W	77	83	82	90	76	64	69	68	59	63	222,4146	-67,5375
M	75	76	69	83	83	70	79	80	77	70	236,4790	-45,7033
W	60	61	56	62	57	73	65	70	81	77	209,5833	-14,6615
M	73	78	76	72	75	84	97	94	85	86	259,1273	-23,1683
W	94	91	85	92	88	80	77	74	73	83	256,8607	-66,0052
W	59	52	53	54	52	74	74	68	78	66	200,7141	-6,1004
W	71	72	76	84	77	79	67	73	74	73	230,8406	-45,8205
W	66	70	67	76	72	80	73	73	79	82	231,2206	-29,8159
M	72	70	71	69	68	61	56	57	55	55	193,4246	-55,2790
W	70	62	79	63	76	74	73	77	77	74	226,7279	-31,6148
M	66	62	56	74	61	68	69	61	62	71	202,7457	-32,8389

Fehlende Daten und Mit Daten verknüpfen

- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert. Fehlende Kategorien werden als „N/A“ umbenannt, wenn sie nach einer Kategorien-Variable aufgeschlüsselt werden.
- **Mit Daten verknüpfen –** Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

Befehl Clusteranalyse

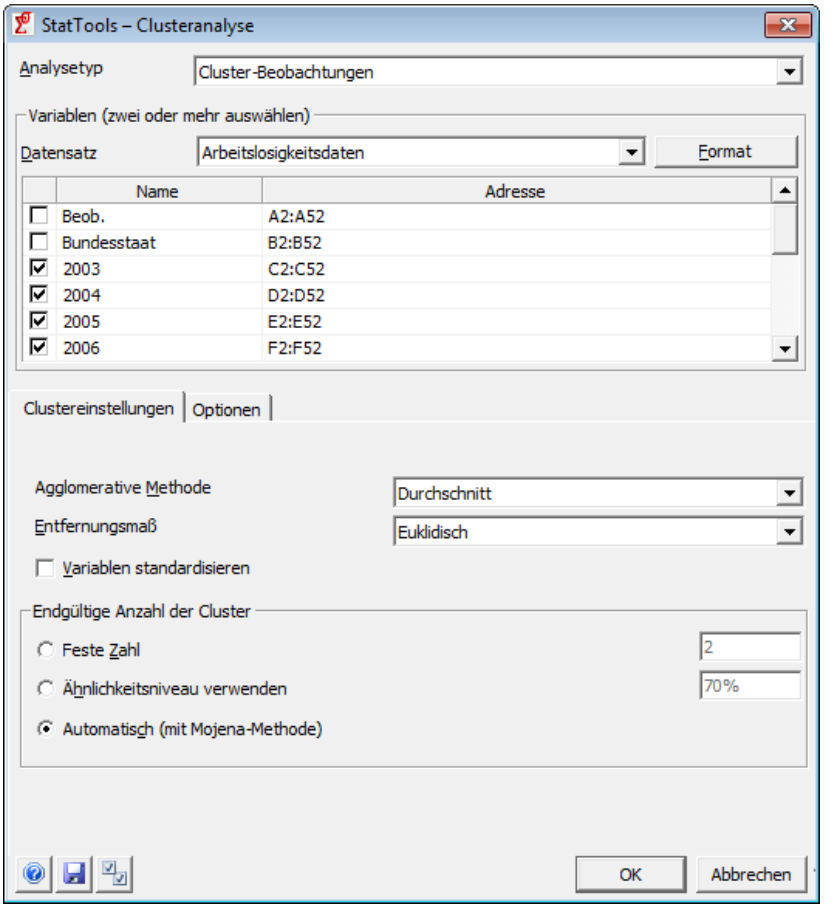
Führt eine Clusteranalyse an einem Datensatz aus

Der Befehl Clusteranalyse sucht nach Mustern in einem Datensatz, um die Beobachtungen oder Variablen in Gruppen ähnlicher Elemente zu klassifizieren. Dieses Verfahren wendet mehrere „agglomerative hierarchische“ Methoden an. Jede dieser Methoden beginnt mit jeder Beobachtung in einem separaten Cluster und kombiniert dann in einer Serie von Schritten die Cluster, bis schlussendlich ein einziger Cluster mit allen Beobachtungen erhalten wird. Durch die Auswahl eines dieser Schritte im Verfahren vor dem Erhalt des letzten Clusters erhalten Sie eine Unterteilung des Datensatzes in Gruppen ähnlicher Elemente.

Die Analyse unterstützt verschiedene agglomerative hierarchische Methoden und Entfernungsmaße. Sie können auch unter drei Optionen zur Ermittlung der endgültigen Anzahl von Clustern auswählen. Das Cluster-Verfahren wird durch ein Diagramm zusammengefasst, das als „Dendrogramm“ bezeichnet wird, sowie durch eine Tabelle mit den einzelnen Schritten des Verfahrens.

Zum Definieren dieser Analyse wird das Dialogfeld Clusteranalyse verwendet:

Dialogfeld Clusteranalyse



Es können zwei Variablen ausgewählt werden. Bei dem ausgewählten Datensatz muss es sich immer um einen entstapelten handeln. Die Variablen können dabei aus verschiedenen Datensätzen stammen.

Oben im Dialogfeld Clusteranalyse hat der Benutzer folgende Wahlmöglichkeiten:

Analysetyp – Wählt die Art der auszuführenden Analyse aus: Cluster-Beobachtungen zum Klassifizieren von Beobachtungen in Gruppen und Cluster-Variablen zum Klassifizieren von Variablen in Gruppen.

Dialogfeld Clusteranalyse – Registerkarte Cluster-einstellungen

Die erste Registerkarte in diesem Dialogfeld (Clustereinstellungen) dient zum Festlegen folgender Optionen:

- **Agglomerative Methode** – Hierarchische Clustering ist ein sequenzieller Vorgang. In jedem Schritt des agglomerativen hierarchischen Ansatzes wird eine Beobachtung oder ein Cluster von Beobachtungen mit einem anderen Cluster zusammengeführt. Um zu entscheiden, welche Cluster-Paare in jedem Schritt zusammengeführt werden sollen, berechnen diese Methoden die Entfernungen zwischen Beobachtungen, verwenden diese zum Berechnen der Entfernungen zwischen Clustern und führen dann die beiden Cluster mit der kleinsten Entfernung zusammen. Verschiedene agglomerative Clustermethoden sind verschiedene Methoden zum Definieren der Entfernung zwischen Clustern:
  - **Einzelne Verknüpfung (nächster Nachbar) –** Hier wird die Entfernung zwischen zwei Clustern A und B als die Mindestentfernung zwischen einem Punkt in A und einem Punkt in B definiert:
    $d(A,B) = min\{d(xi,yj), für\ xi\ in\ A\ und\ yj\ in\ B\}$ wobei $d(xi,yj)$ die Entfernung zwischen Vektoren xi und yj ist.
  - **Komplette Verknüpfung (Entferntester Nachbar) –** Hier wird die Entfernung zwischen zwei Clustern A und B als die Maximalentfernung zwischen einem Punkt in A und einem Punkt in B definiert:
    $d(A,B) = max\{d(xi,yj), für\ xi\ in\ A\ und\ yj\ in\ B\}$ wobei $d(xi,yj)$ die Entfernung zwischen Vektoren xi und yj ist.
  - **Durchschnittsverknüpfung –** Hier wird die Entfernung zwischen zwei Clustern A und B als der Durchschnitt der Entfernungen nA*nB zwischen Punkten nA in A und Punkten nB in B definiert:
    $d(A,B) = \sum\sum d(xi,yj) / (nA*nB)$ wobei die Summe insgesamt xi in A und alle yj in B ist.
  - **Flächenmittelpunkt –** Hier wird die Entfernung zwischen zwei Clustern A und B als die Entfernung zwischen den mittleren Vektoren (häufig als Flächenmittelpunkte bezeichnet) der beiden Cluster definiert:
    $d(A,B) = d(\bar{x}A, \bar{y}B)$
    wobei $\bar{x}A$ und $\bar{y}B$ die mittleren Vektoren für die Beobachtungsvektoren in A bzw. die Beobachtungsvektoren in B sind.

- **Median –** Diese Option ist der Option Flächenmittelpunkt ähnlich: Der Unterschied besteht darin, dass der Flächenmittelpunkt für den neuen Cluster als der Medianwert (Mittelpunkt) zwischen den Flächenmittelpunkten der alten Gruppen berechnet wird *centroid(A,B) = (*$\bar{x}A$ + $\bar{y}B$*) / 2*

  wobei $\overline{\mathbf{x}}_A$ und $\overline{\mathbf{y}}_B$ die „Flächenmittelpunkte" für Cluster A bzw. Cluster B sind. (Es ist zu beachten, dass dieser Medianwert nicht der in der Statistik übliche Medianwert ist.) Die Entfernung zwischen Clustern ist die Entfernung zwischen den auf diese Weise berechneten Cluster-Flächenmittelpunkten.

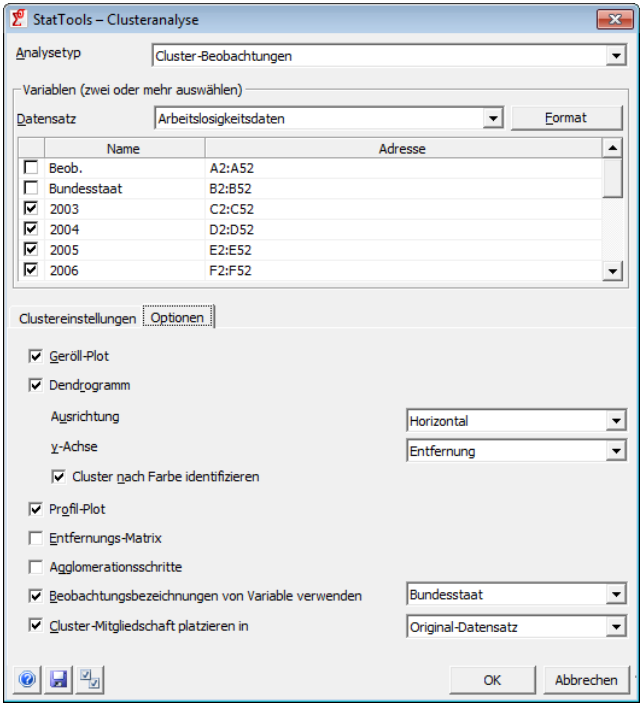
- **Methode von Ward –** Diese Methode wird genauer als Minimalstreuungsmethode von Ward bezeichnet, da sie das Kriterium der Minimalstreuung zur Auswahl des Cluster-Paars verwendet, das in jedem Schritt zusammengeführt wird.

- **Entfernungsmaß –** Legt das Entfernungsmaß fest, das für die ausgewählte agglomerative Methode verwendet werden soll. Es sind folgende Optionen verfügbar:
  - **Euklidisch**
  - **Quadrat-Euklidisch**
  - **Mahalanobis**
  - **Manhattan** (auch als „Stadtblock"-Entfernung bezeichnet)
  - **Korrelation**, wenn Variablen geclustert werden
  - **Absolute Korrelation**, wenn Variablen geclustert werden
- **Variablen standardisieren –** Der Mittelwert wird von jeder Variable abgezogen und das Ergebnis dann durch die Standardabweichung dividiert, bevor die Entfernungsmatrix berechnet wird. Verwenden Sie diese Option, wenn die Maßeinheiten verschieden sind und die Auswirkungen dieser Unterschiede minimiert werden sollen.
- **Endgültige Anzahl der Cluster –** Legt die endgültige Anzahl der zu erstellenden Cluster fest. Es sind folgende Optionen verfügbar:
  - **Feste Zahl**

- **Ähnlichkeitsniveau verwenden –** Das Ähnlichkeitsniveau zwischen zwei Clustern A und B ist wie folgt definiert:
  $s(A,B)=100(1-d(A,B))/dmax$ wobei $d(A,B)$ die Entfernung zwischen Clustern A und B ist und dmax der Maximalwert in der ursprünglichen Entfernungsmatrix D ist.

- **Automatisch (mit Mojena-Methode) –** Die Anzahl der ausgewählten Cluster entspricht der ersten Stufe im Dendrogramm, das die folgende Bedingung erfüllt:
  $aj > \bar{a} + ksa$ *für* $j=1,2,...,n$
  wobei α1,α2, ... ,αn die Entfernungswerte für Stufen mit n, n-1,…,1 Clustern, $\bar{\alpha}$ und sα der Mittelwert bzw. die Standardabweichung der einzelnen α-Werte und k eine Konstante sind. Milligan und Cooper empfehlen k = 1,25, basierend auf einer Simulationsstudie („An examination of procedures for determining the number of clusters in a data set", Psychometrika, Juni 1985, Jahrgang 50, Ausgabe 2, Seite 159-179).

Dialogfeld Clusteranalyse – Registerkarte Optionen



Mit den Optionen in diesem Dialogfeld werden Einstellungen für die erzeugten Diagramme, Tabellen und Variablen festgelegt. Es sind folgende Optionen verfügbar:

- **Geröll-Plot –** Trägt die Anzahl der Cluster auf der x-Achse (beginnend mit der Ein-Cluster-Lösung am Ursprung) gegen die Entfernung auf, bei der die Objekte oder Cluster auf der y-Achse kombiniert werden.
- **Dendrogramm –** Zeichnet ein Baumstruktur-Diagramm, das sowohl den Zusammenhang zwischen Clustern und Subclustern als auch die Reihenfolge, in der die Cluster zusammengeführt wurden, zeigt. Für dieses Diagramm sind mehrere Einstellungen verfügbar:
- **Ausrichtung** – Horizontal oder vertikal.
- **Y-Achse** – Legt die Maßeinheit für die y-Achse fest.
- **Cluster nach Farbe identifizieren** – Die Höchstanzahl der verfügbaren Farben ist 15.
- **Profil-Plot –** Das ist eine Serie von farbigen Linien, wobei jede Linie einen Cluster darstellt. Für jeden Cluster wird der Mittelwert einer jeden Variable aufgetragen, und die Punkte werden durch eine farbige Linie verbunden, die diesen Cluster darstellt. Nahe aneinander liegende Linien geben ähnliche Cluster an; weit voneinander entfernte Linien geben unähnliche Cluster an.
- **Entfernungs-Matrix –** Zeigt die anfängliche Entfernungs-Matrix.
- **Agglomerationsschritte –** Zeigt die Details der Agglomerationsschritte.
- **Beobachtungsbezeichnungen von Variable verwenden –** Wählt eine Variable mit Bezeichnungen für Beobachtungen aus. Diese Variable sollte keine sich wiederholenden Werte haben. Sie wird im Dendrogramm und in der Tabelle mit den Agglomerationsschritten verwendet.
- **Cluster-Mitgliedschaft platzieren in –** Setzt eine Variable ein, damit erkenntlich ist, zu welchem Cluster die einzelnen Beobachtungen gehören. Wenn dies berücksichtigt wird, können Sie festlegen, wo diese Variablen platziert werden sollen: im ursprünglichen Datensatz oder in einem neuen.

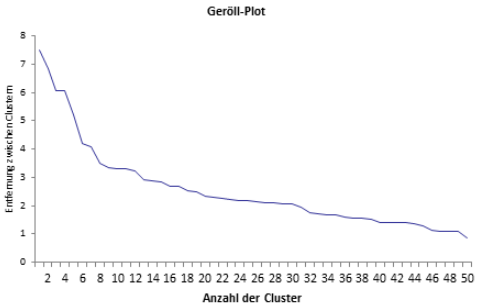
Clusteranalyse-Bericht

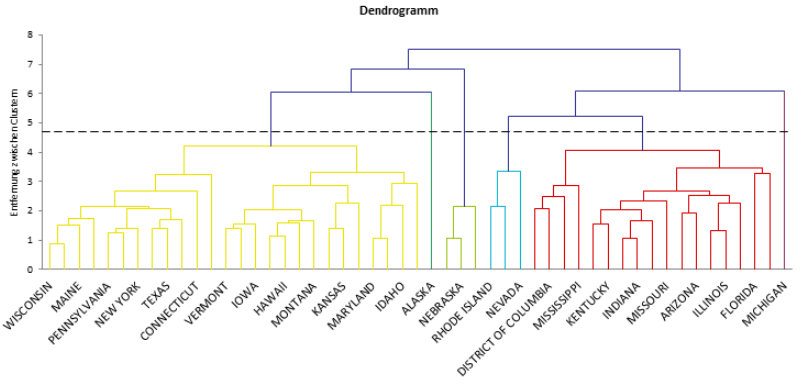
StatTools-Bericht

Analyse: Clusteranalyse
Ausgeführt durch: Palisade
Datum: Donnerstag, 18. Juni 2015
Aktualisierung: Statisch

Clusteranalyse

Analysetyp	Cluster-Beobachtungen
Agglomerative Methode	Durchschnitt
Entfernungsmaß	Euklidisch
Standardisierte Variabl	Nein
Endgültige Anzahl der	Automatisch (mit Mojena-Methode)
Anzahl der Cluster	6
Ähnlichkeitsniveau	77,27%





Datensatz mit Cluster-Mitgliedschaft

Wenn die Option Cluster-Mitgliedschaft platzieren in aktiviert ist, wird eine Spalte mit der Gruppenmitgliedschaft rechts des ausgewählten Datensatzes eingetragen.

Beob.	Bundesstaat	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Cluster
1	ALABAMA	5,6	5	3,8	3,5	3,4	5	9,8	9,3	8,7	7,3	4
2	ALASKA	7,4	7,4	6,9	6,5	6,1	6,4	7,7	8	7,6	7	2
3	ARIZONA	6,3	5	4,7	4,1	3,7	6	9,8	10,4	9,4	8,3	4
4	ARKANSAS	5,5	5,6	5,1	5,3	5,3	5,4	7,5	7,9	7,9	7,3	3
5	CALIFORNIA	6,8	6,2	5,4	4,9	5,4	7,2	11,3	12,4	11,8	10,5	5
6	COLORADO	5,7	5,6	5,1	4,3	3,8	4,8	8,1	9	8,6	8	3
7	CONNECTICUT	4,5	4,9	4,9	4,4	4,6	5,6	8,2	9,3	8,9	8,4	3
8	DELAWARE	4,1	3,9	4	3,5	3,5	4,9	7,9	8	7,4	7,1	3
9	DISTRICT OF COLUMBIA	7	7,5	6,5	5,7	5,5	6,6	9,7	10	10,1	8,9	4
10	FLORIDA	5,8	4,7	3,8	3,3	4	6,3	10,4	11,3	10,3	8,6	4
11	GEORGIA	5,2	4,7	5,2	4,7	4,6	6,3	9,8	10,2	9,9	9	4
12	HAWAII	4,2	3,2	2,8	2,5	2,7	4,1	6,9	6,8	6,5	5,8	3

Fehlende Daten und Mit Daten verknüpfen

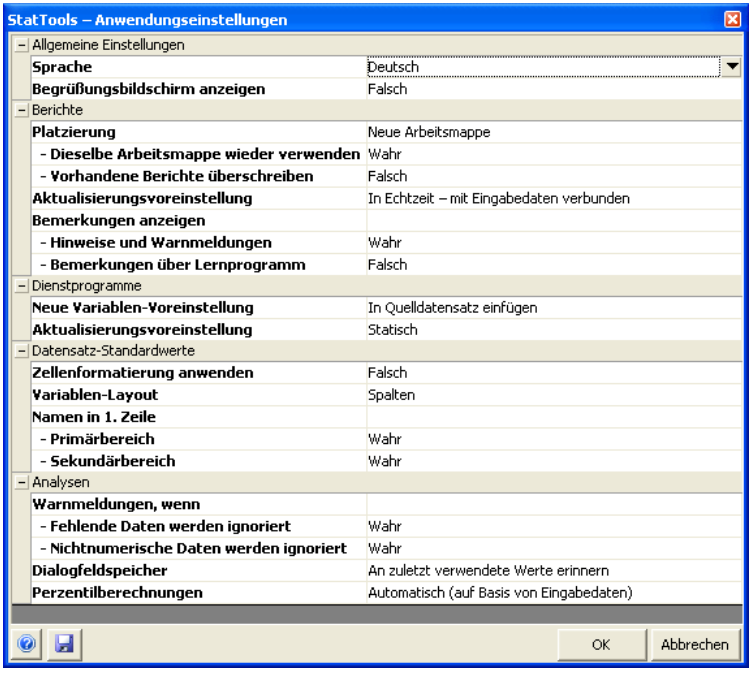
- **Fehlende Daten –** Fehlende Daten stellen kein Problem dar. In den ausgewählten Variablen werden alle Zeilen mit fehlenden Daten einfach ignoriert.
- **Mit Daten verknüpfen –** Es ist keine Verknüpfung mit den Originaldaten vorhanden. Falls die Daten sich ändern, muss die Analyse erneut ausgeführt werden.

Menü *Dienstprogramme*

Befehl *Anwendungseinstellungen*

Legt in StatTools die Einstellungen für Berichte, Diagramme, Dienstprogramme, Datensätze und Analysen fest

Dieser Befehl ermöglicht Ihnen, in StatTools die allgemeinen Einstellungen für Berichte, Diagramme, Dienstprogramme, Datensätze und Analysen anzugeben. Diese Einstellungen werden dann auf alle Analysen und Datensätze angewendet. Andere analysenbezogene Einstellungen können jeweils im Dialogfeld für die einzelnen Analysen definiert werden.



Unter **Berichte** sind die Optionen für durch StatTools-Analysen erstellte Berichte und Diagramme angegeben. Es handelt sich dabei um folgende Optionen:

- **Platzierung** – wählt aus, wo in Excel die neuen Berichte und Diagramme erstellt werden sollen:
  - **Aktive Arbeitsmappe** – was bedeutet, dass in dieser Arbeitsmappe ein neues Arbeitsblatt für jeden Bericht erstellt wird
  - **Neue Arbeitsmappe** – was bedeutet, dass nötigenfalls eine neue Berichtsarbeitsmappe in StatTools erstellt und dann jeder Bericht darin auf einem Arbeitsblatt platzieret wird.
  - **Nach zuletzt verwendeter Spalte im aktiven Arbeitsblatt** – was bedeutet, dass in StatTools jeder Bericht rechts der zuletzt verwendeten Spalte auf einem aktiven Arbeitsblatt platziert wird.
  - **Abfrage nach Startzelle** – was bedeutet, dass Sie nach Ausführung einer Analyse eine Zelle auswählen können, in der dann die linke obere Kante des Berichts oder Diagramms platziert wird.
  - **Dieselbe neue Arbeitsmappe erneut verwenden** – was bedeutet, dass die neu erstellte Arbeitsmappe für alle Berichte zu verwenden ist.
- **Aktualisierungsvoreinstellung** – legt fest, wie die Ergebnisse aktualisiert werden sollen, wenn sich die Daten in den Variablen ändern. Für diese Ergebnisaktualisierung stehen folgende Optionen zur Verfügung:
  - **Echtzeit – Werte ändern sich gemäß Eingabedaten** – was bedeutet, dass Berichte bei Änderung der Eingabedaten automatisch aktualisiert werden.
  - **Statisch – feststehende Werte** – was bedeutet, dass Berichte bei sich ändernden Eingabedaten unverändert bleiben. Die Statistiken werden nicht aktualisiert und basieren stets auf den Werten der Daten, die bei Ausführung des Vorgangs eingegeben wurden.

Um in StatTools Echtzeit-Ergebnisse zu erhalten, müssen Excel-Formeln und benutzerdefinierte StatTools-Funktionen verwendet werden. Durch die Formel

=StatMean('Vertrauensintervall.xls'!Paar)

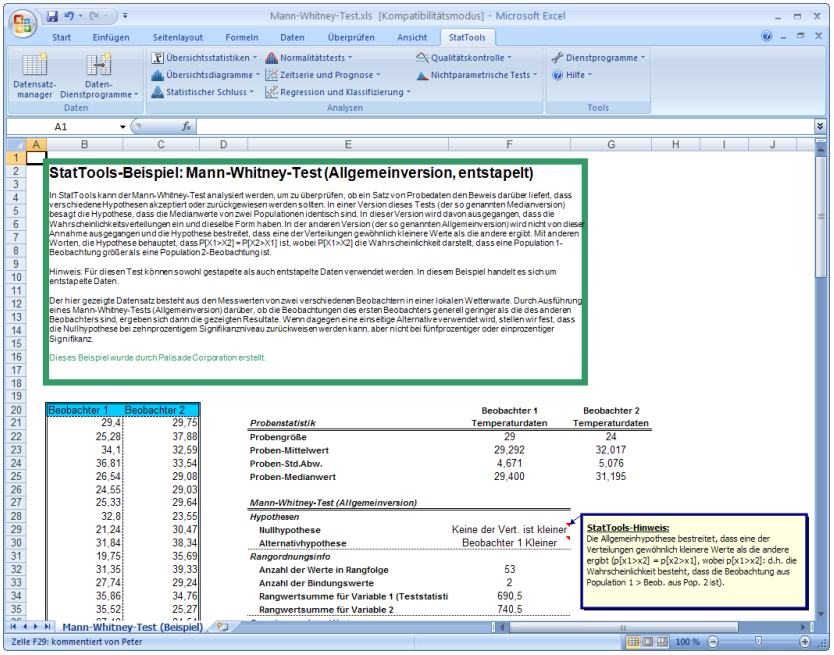
wird beispielsweise der Mittelwert der Variable *Paar* berechnet. In dieser Variable werden Daten aus dem Excel-Bereich *Paar* verwendet, der in der Arbeitsmappe *Vertrauensintervall.xls* zu finden ist. Mit den sich ändernden Daten im Bereich *Paar* werden die durch die Funktion *StatMean* zurückgegebenen Werte automatisch entsprechend aktualisiert.

Durch StatTools-Vorgänge erstellte Berichte und Diagramme können in Echtzeit aktualisiert werden. Es bestehen jedoch folgende Ausnahmen:

1) **Regression**
2) **Diskriminanzanalyse**
3) **Logistische Regression**
4) **Prognose**

Diese Vorgänge machen lange Neuberechnungen erforderlich, wodurch Excel während einer Echtzeit-Aktualisierung erheblich langsamer reagieren würde.

- **Bemerkungen anzeigen** – legt fest, welche Kategorien von StatTools-Meldungen in Berichte mit einbezogen werden sollen. Hinweise, Warnmeldungen und Lernprogrammhinweise können in Berichtszellen als Popup-Hinweise angezeigt werden, wie hier gezeigt ist:



Dienstprogramm-einstellungen

Unter **Dienstprogramme s**ind die Optionen für neue durch StatTools-Dienstprogramme zu erstellende Variablen angegeben. Diese Optionen sind im Menü **Dienstprogramme** zu finden. Über diese Optionen können neue Variablen erstellt werden, indem bereits bestehende Variablen transformiert, kombiniert oder sonst irgendwie verarbeitet werden.

- **Neue Variablen-Voreinstellungen** – legt fest, wo die neuen durch Dienstprogramme erstellten Variablen platziert werden sollen. Dafür sind folgende Optionen verfügbar:
  - **In Quelldatensatz einfügen** – was bedeutet, dass jede neu erstellte Variable rechts (oder unten) in dem Datensatz eingefügt wird, der die Originalvariable enthält.
  - **Neuen Datensatz erstellen** – was bedeutet, dass die neuen Variablen in einem neuen Datensatz platziert werden.

Es gibt jedoch Fälle, in denen die ausgewählte neue Variablen-Voreinstellung nicht befolgt wird. Das passiert beispielsweise bei **Stapeln und Entstapeln** (wobei neue Variablen immer in einem neuen Datensatz platziert werden) oder wenn die Originaldaten aus einem Mehrfachbereichsdatensatz stammen (in welchem Fall neue Variablen nur in den Quelldatensatz eingefügt werden können).

- **Aktualisierungsvoreinstellung** – legt fest, wie die Werte für neue durch ein Dienstprogramm erstellte Variablen aktualisiert werden sollen, wenn sich die Daten in der Originalvariable ändern. Hier stehen folgende Optionen zur Verfügung:
  - **Echtzeit – Werte ändern sich gemäß Eingabedaten** – was bedeutet, dass neue Variablenwerte bei Änderung der Eingabedaten automatisch aktualisiert werden.
  - **Statisch – feststehende Werte** – was bedeutet, dass neue Variablenwerte bei sich ändernden Eingabedaten unverändert bleiben. Neue Variablenwerte werden nicht aktualisiert und basieren stets auf den Datenwerten, die bei Ausführung des Vorgangs eingegeben wurden.

  Es gibt jedoch Fälle, in denen die ausgewählte Aktualisierungsvoreinstellung nicht befolgt und für neue Variablen stets die Präferenz **Statisch – feststehende Werte** verwendet wird. Das ist beispielsweise bei **Stapeln und Entstapeln** sowie **Zufallsprobewerte** der Fall. Bei diesen Dienstprogrammen kann die Echtzeit-Aktualisierung nicht verwendet werden.

Datensatz-einstellungen

Über **Datensatz-Standardwerte** können die Optionen für neue Datensätze angegeben werden, die mithilfe des Befehls **Datensatzmanager** erstellt wurden. Diese Einstellungen sind lediglich Standardwerte, die beim Erstellen eines neuen Datensatzes angezeigt werden. Sie können diese Einstellungen nötigenfalls im Dialogfeld **Datensatzmanager** ändern.

- **Zellenformatierung anwenden** – bestimmt, dass der Datensatz durch StatTools formatiert werden soll.
- **Layout** – wählt das standardmäßige Layout für die Variable aus (zeilenweise oder spaltenweise).
- **Namen in erster Spalte/Zeile (Primärbereich)** – legt fest, dass die Namen für den ersten definierten Bereich in die erste Spalte oder Zeile eingegeben werden sollen.
- **Namen in erster Spalte/Zeile (Sekundärbereich) –** legt fest, dass die Namen für einen Mehrfachbereichsdatensatz in die erste Spalte oder Zeile des zweiten und aller nachfolgenden Bereiche eingegeben werden sollen.

Analysen-einstellungen

Über die **Analysen** werden die Standardwerte angegeben, die in den zum Definieren von Analysen verwendeten Dialogfeldern angezeigt werden sollen. Dies sind die Einträge, die Sie sehen, wenn Sie beim Definieren einer Analyse das Dialogfeld erstmalig anzeigen lassen. Nötigenfalls können diese Einträge in den einzelnen Dialogfeldern noch geändert werden.

- **Warnmeldungen** – legt fest, ob Warnmeldungen beim Ausführen einer Analyse angezeigt werden sollen, wenn StatTools feststellt, dass in einer Variable irgendwelche Daten fehlen oder nichtnumerische Daten vorhanden sind.

- **Dialogfeldspeicher –** bestimmt die Standardeinträge, die in den zum Definieren von Analysen verwendeten Dialogfeldern angezeigt werden sollen. Hier können folgende Optionen gewählt werden:
  - **An zuletzt verwendete Werte erinnern (bei Arbeitsmappe) –** was bedeutet, dass im angezeigten Dialogfeld die Einträge zu sehen sind, die schon das letzte Mal zu sehen waren, als dieses Dialogfeld für die aktive Arbeitsmappe angezeigt wurde. Wenn das Dialogfeld dagegen erstmalig angezeigt wird, sind die gespeicherten Systemstandardeinstellungen für die Analyse zu sehen.
  - **Immer Systemstandardwerte verwenden –** was bedeutet, dass im angezeigten Dialogfeld immer die für die Analyse gespeicherten Systemstandardeinstellungen zu sehen sind.

- **Perzentilberechnungen –** wählt die Methode aus, die zum Berechnen von Perzentilen verwendet werden soll. Je nach Art der Daten können durch verschiedene Methoden bessere Ergebnisse erzielt werden. Hier sind die verfügbaren Methoden (und die Datentypen, für die sie geeignet sind):

  1) Automatisch (auf Basis von Eingabedaten)
  2) Mit asymmetrischen Endpunkten interpoliert (kontinuierlich)
  3) Mit symmetrischen Endpunkten interpoliert (kontinuierlich)
  4) Excel-Perzentilfunktion (kontinuierlich)
  5) Nächstbeste Beobachtung (diskontinuierlich)
  6) Erfahrungsgemäße Verteilungsfunktion (diskontinuierlich)
  7) Erfahrungsgemäße Verteilungsfunktion mit Mittelwertbildung (diskontinuierlich)

Befehl *Datensätze löschen*

Löscht in StatTools die Datensätze in der aktiven Arbeitsmappe

Mithilfe dieses Befehls können in der aktiven Arbeitsmappe alle definierten Datensätze gelöscht werden. Die eigentlichen Daten in Excel werden dadurch nicht gelöscht, sondern nur die Definition der Datensätze.

Befehl *Dialogfeldspeicher löschen*

Löscht alle gespeicherten Einträge in den Dialogfeldern für die Analysen

Über diesen Befehl können alle gespeicherten Analysen-Dialogfeldeinträge gelöscht werden. In nachfolgend angezeigten Dialogfeldern sind dann anfänglich die gespeicherten Systemstandardeinstellungen für jede Analyse zu sehen.

Befehl *StatTools-Add-In entladen*

Entlädt das StatTools-Add-In

Über diesen Befehl kann StatTools entladen und können alle StatTools-Fenster geschlossen werden.

Hilfemenü

StatTools-Hilfe

Öffnet die Online-Hilfedatei für StatTools

Über den Befehl **StatTools-Hilfe** aus dem Menü **„?“** können Sie die StatTools-Haupthilfedatei öffnen. In dieser Datei werden alle StatTools-Funktionen und -Befehle beschrieben.

Befehl *Online-Handbuch*

Öffnet das Online-Handbuch für StatTools

Durch den Befehl **Online-Handbuch** aus dem Menü **„?“** kann dieses Handbuch im PDF-Format geöffnet werden. Dafür muss allerdings auf dem Rechner das Programm **Adobe Acrobat Reader** installiert sein.

Befehl *Lizenzaktivierung*

Zeigt die Lizenzierungsinformationen für StatTools an und ermöglicht die Lizenzierung von Demo-Versionen

Über den Befehl **Lizenzaktivierung** aus dem Menü **?** kann das Dialogfeld **Lizenzaktivierung** angezeigt werden, in dem die Versions- und Lizenzierungsinformationen für Ihr StatTools-Programm zu finden sind. Sie können dieses Dialogfeld auch dazu verwenden, eine Demo-Version von StatTools in ein lizenziertes Programm konvertieren zu lassen.

Weitere Informationen über die Lizenzierung von StatTools sind in diesem Handbuch in **Kapitel 1: Erste Schritte** zu finden.

Befehl *Info über*

Zeigt die Versions- und Copyright-Informationen über StatTools an

Über den Befehl **Info über** aus dem Hilfemenü können Sie das Dialogfeld **Info über** anzeigen lassen, in dem die Versions- und Copyright-Informationen über StatTools aufgeführt sind.

Referenz: StatTools-Funktionen

Einführung

In StatTools werden benutzerdefinierte Arbeitsblattfunktionen verwendet, um berechnete Statistiken an die Excel-Formeln zurückzugeben. Durch diese Funktionen ist es möglich:

1) Statistikberechnungen in Arbeitsblattformeln einzubetten, und zwar genauso, wie das bei standardmäßigen Excel-Funktionen der Fall ist

2) Statistiken in Echtzeit zu aktualisieren, d.h., die Ergebnisse ändern sich, sobald sich die Originaldaten ändern

Wenn Sie sich in einem StatTools-Bericht die Formeln in den Zellen ansehen, können Sie die Arbeitsblattfunktionen erkennen. Alle StatTools-Funktionen beginnen mit dem Präfix „Stat“, wie z.B. **StatMean()** oder **StatStdDev().** Der Einfachheit halber sind alle StatTools-Funktionen im Excel-Dialogfeld **Funktion einfügen** aufgeführt.

StatTools-Funktionen gegenüber Excel-Funktionen

In einigen Fällen ersetzt StatTools die in Excel integrierte Statistik durch seine eigenen robusten und schnellen Berechnungen. Die Genauigkeit der in Excel verwendeten Statistiken ist schon oft in Frage gestellt worden und daher werden diese in StatTools überhaupt nicht eingesetzt! Sogar die in Excel vorhandenen Arbeitsblattstatistikfunktionen, wie z.B. *STABW()*, sind durch neue, robuste StatTools-Versionen, wie z.B. *StatSTDEV()*, ersetzt worden. Die in StatTools verwendeten Statistikberechnungen sind äußerst genau und leistungsoptimiert durch Verwendung von C++ DLLs anstelle von Makro-Berechnungen.

Im Gegensatz zu den integrierten Excel-Funktionen unterstützen die StatTools-Funktionen die Verwendung von gestapelten Daten. Durch die Funktion **StatDestack** werden automatisch Daten aus einem gestapelten Datensatz für die von Ihnen angegebene Kategorie entstapelt. Anschließend werden diese Daten dann an eine StatTools-Statistikfunktion weitergegeben, um analysiert zu werden.

Über StatTools-Funktionen können auch Daten analysiert werden, die sich auf verschiedenen Arbeitsblättern befinden. Durch Mehrfachblatt-Datensätze sind mehr als 65535 Punkte pro Variable möglich. Diese Punkte werden im Dialogfeld **Datensatzmanager** eingegeben, und zwar über die Schaltfläche **Mehrfach**.

Verteilungsfunktionen

StatTools enthält einen Satz von Verteilungsfunktionen (wie z.B. **StatBionomial**), durch die die in Excel integrierten Verteilungsfunktionen (wie z.B. **BinomVert**) ersetzt werden. Im Gegensatz zu den in Excel enthaltenen Verteilungsfunktionen kann durch die Verteilungsfunktionen in StatTools eine Anzahl von verschiedenen Werten aus einer Wahrscheinlichkeitsverteilung zurückgegeben werden. Der zurückzugebende Wert wird über das Argument **Statistik** angegeben, bei dem es sich um das zweitletzte Argument in der Funktion handelt. Dieses Argument kann den Wert 1 bis 12 haben oder auch aus einer Zeichenfolge bestehen, durch die auf die Statistik hingewiesen wird, die für die eingegebene Verteilung abgerufen werden soll.

Liste der möglichen Statistiken, die zurückgegeben werden können

Wert oder Zeichenfolge	Zurückgegebene Statistik
1 oder „Mittelwert“	Mittelwert
2 oder „Std.Abw.“	Standardabweichung:
3 oder „Varianz“	Varianz
4 oder „Schiefe“	Schiefe
5 oder „Wölbung“	Wölbung
6 oder „Modus“	Modus:
7 oder „disk. Mittelw.“	diskontinuierlicher Mittelwert (oder der Wert, der dem wahren Mittelwert (der wirklich auftreten könnte) am nächsten kommt)
8 oder „x zu y“	x zu y (der y-Wert der Verteilung für den eingegebenen x-Wert)
9 oder „x zu p“	x zu p (der p-Wert der Verteilung für den eingegebenen x-Wert)
10 oder „p zu x“	p zu x (der x-Wert der Verteilung für den eingegebenen p-Wert)
11 oder „x zu q“	x zu q (der q-Wert der Verteilung für den eingegebenen x-Wert)
12 oder „q zu x“	q zu x (der x-Wert der Verteilung für den eingegebenen q-Wert)

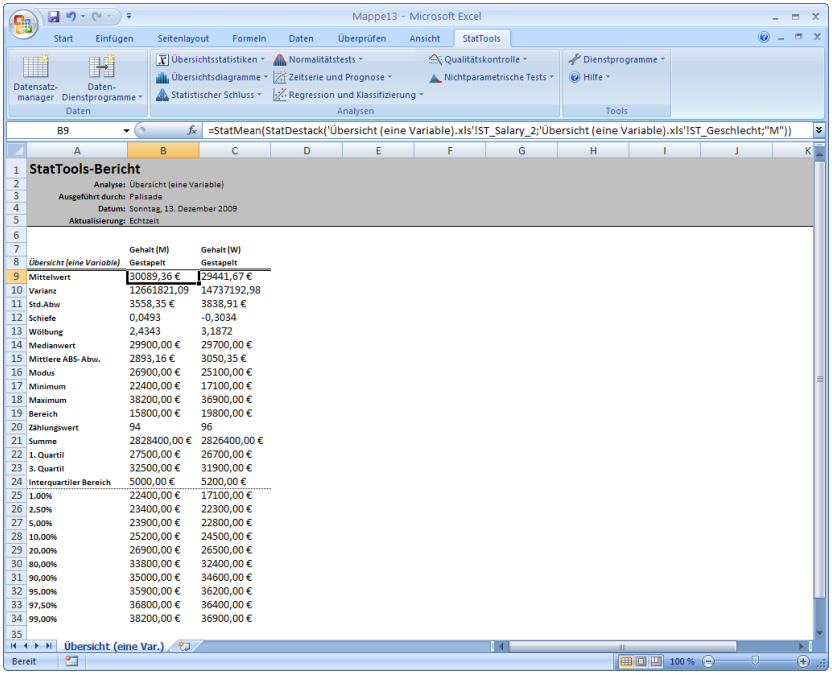
Durch die StatTools-Verteilungsfunktion

StatNormal(10;1;"x zu p"; 9,5)

wird beispielsweise der p-Wert zurückgegeben, der mit dem x-Wert von 9,5 verbunden ist, und zwar in einer Normalverteilung mit einem Mittelwert von 10 und einer Standardabweichung von 1.

Echtzeit-Berichte

StatTools verwendet benutzerdefinierte Funktionen, um die Ergebnisse möglichst in Echtzeit anzuzeigen. Mit anderen Worten, Berichte werden möglichst durch Formeln mit den Originaldaten verknüpft. Angenommen, Sie arbeiten mit der Variable *Gewicht* und möchten Übersichtsfaktoren über das *Gewicht*, wie z.B. Mittelwert und Standardabweichung, haben. Durch den Vorgang **Übersichtsstatistik** erhält der Gewichtsbereich den Namen *Gewicht* und werden dann folgende Formeln in die Ausgabezellen eingegeben: **=StatMean(Gewicht)** und **=StatStdDev(Gewicht).** Bei StatMean und StatStdDev handelt es sich um in StatTools integrierte Funktionen, die zum Berechnen des Mittelwerts und der Standardabweichung verwendet werden. Dadurch werden die standardmäßigen für diese Statistiken bestimmten Excel-Funktionen ersetzt. Durch diese Funktionen ändern sich automatisch die Ergebnisse, sobald die Daten irgendwie verändert werden. Dadurch braucht der Vorgang dann nicht erneut ausgeführt werden.



Mitunter ist es jedoch nicht angebracht, die Ergebnisse mit den Daten zu verknüpfen. Ein gutes Beispiel dafür ist die Regression. StatTools gibt Ihnen nicht die Formeln, die zum Erstellen der Regressionsausgabe verwendet werden, sondern nur die numerischen Ergebnisse. In solchen Fällen müssen Sie also bei Datenänderung die Vorgänge erneut ausführen.

Sie können unter dem Befehl **Berichtseinstellungen** die Option **Statisch** verwenden, um die Berichte nicht mit den Daten verknüpfen zu lassen. Das kann von Nutzen sein, wenn in Excel die Neuberechnung bei Datenänderung zu lange dauert.

Referenz: Liste der Statistikfunktionen

Tabelle der verfügbaren Funktionen

In dieser Tabelle sind die anpassbaren Funktionen aufgeführt, die durch StatTools dem Excel-Programm hinzugefügt werden. Bei Verwendung werden alle Funktionen mit dem Präfix **Stat** versehen.

Funktion	Rückgabe
AutoCorrelation(*Daten; Verzöger.*)	Berechnet die Autokorrelation für die Daten in einem Datensatz
AveDev(*Daten1; Daten2; ... DatenN*)	Berechnet die durchschnittliche absolute Abweichung der Daten von ihrem Mittelwert. Die Argumente können aus Zahlen, Matrizen oder Bereichen bestehen.
BINOMIAL(*N; P; Statistik; Wert*)	Berechnet die Statistik für die eingegebene Binomial-Verteilung
CATEGORYINDICIES (*Bereich; Kategoriename*)	Ruft die Zellenindizes für die angegebene Kategorie ab
CATEGORYNAMES(*Bereich*)	Ruft die Kategorienamen in einem Bereich ab
CATEGORYOCCURRENCECOUNT (*Bereich; Kategoriename*)	Berechnet die Zellenanzahl in einem Bereich der angegebenen Kategorie
ChiSQ (*Freiheitsgrad; Statistik; Wert*)	Berechnet die Statistik für die eingegebene schiefe Chi-Quadrat-Verteilung
CorrelationCoeff(*Daten1; Daten2; Flag*)	Berechnet den Korrelationskoeffizienten zwischen zwei Datensätzen. Es kann der lineare (Pearson) oder Rangkorrelations (Spearman)-Koeffizient berechnet werden.
COUNT(*Daten1; Daten2; ... DatenN*)	"Berechnet die Anzahl der Elemente in den betreffenden Argumenten, die aus Zahlen, Matrizen oder Bereichen bestehen können
COUNTCATEGORIES(*Bereich*)	Zählt die Anzahl der Kategorien in einem Bereich
COUNTCELLSBYTYPE (*Bereich; Typ*)	Berechnet die Anzahl der Zellen eines bestimmten Typs in einem Bereich
COUNTRANGE(*Bereich; Minimalwert; Maximalwert; einschl.Min.; einschl.Max.*)	Berechnet die Anzahl der Werte im Bereich, die zwischen den Minimal- und den Maximalwert fallen
Covariance(*Daten1; Daten2; ... DatenN*)	Berechnet die Werteproben-Kovarianz zwischen 2 Datensätzen
COVARIANCEP (*Daten1; Daten2; ... DatenN*)	Berechnet die Populations-Kovarianz zwischen 2 Datensätzen. Falls irgendwelche Zahlen fehlen, wird ein Leerwert zurückgegeben.
Destack(*Datenbereich; Kategoriebereich1; Kategoriebereich2; Kategorie2*)	Extrahiert Daten einer bestimmten Kategorie, und zwar aus gestapelten Daten

Funktion	Rückgabe
DurbinWatson(*Daten*)	Berechnet die Durbin-Watson-Statistik für die Werte in einem Datensatz
F (*Freiheitsgrad1; Freiheitsgrad2; Statistik; Wert*)	Berechnet die Statistik für die eingegebene F-Verteilung von 2 Datensätzen
GETCELLVALUES(*Bereich*)	Ruft die Werte für alle Zellen eines bestimmten Typs im Bereich ab
Kurtosis(*Daten1; Daten2; ... DatenN*)	Berechnet die Werteprobenwölbung der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
KurtosisP(*Daten1; Daten2; ... DatenN*)	Berechnet die Populationswölbung der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
Ln(*x*)	Berechnet den natürlichen Logarithmus einer positiven, reellen Zahl
Max(*Daten1; Daten2; ... DatenN*)	Berechnet das Maximum der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
MEAN(*Daten1; Daten2; ... DatenN*)	Berechnet den arithmetischen Mittelwert (Durchschnitt) der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereich bestehen können
MEANABS(*Daten1; Daten2; ... DatenN*)	Berechnet den arithmetischen Mittelwert (Durchschnitt) der absoluten Werte der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereich bestehen können
MEDIAN (*Daten; diskont.Flag*)	Berechnet den Medianwert eines Datensatzes
MIN(*Daten1; Daten2; ... DatenN*)	Berechnet das Minimum der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
NORMAL (*Mittelwert; Std.Abw.; Statistik; Wert*)	Berechnet die Statistik für die eingegebene normale (Gaußsche) Verteilung
PairCount(*Daten1; Daten2; ... DatenN*)	Zählt die Anzahl der Zellenpaare, für die jede Zelle im Paar numerisch ist
PAIRMEAN (*Daten1; Daten2; ... DatenN*)	Berechnet den Mittelwert der Differenzen zwischen den Zellpaaren
PairMedIAN(*Daten1; Daten2; ... DatenN*)	Berechnet den Medianwert der Differenzen zwischen den Zellpaaren
PAIRSTDDEV (*Daten1; Daten2; ... DatenN*)	Berechnet die Werteproben-Standardabweichung der Differenzen zwischen den Zellpaaren
PerCentile(*Daten; p; diskont.Flag*)	Berechnet das p.-Perzentil eines Datensatzes
PRODUCT (*Daten1; Daten2; ... DatenN*; …DatenN)	Berechnet das Produkt der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können

Funktion	Rückgabe
Quartile(*Daten; q; diskont.Flag*)	Berechnet das angegebene Quartil eines Datensatzes
RAND()	Gibt eine Zufallszahl im Bereich von 0 – 1 zurück
RANGE (*Daten1; Daten2; ... DatenN*)	Berechnet den Bereich (Maximum – Minimum) der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
RunsTest(*Daten; Abstreich*)	Berechnet die Ausführzählungs-Statistik für die Werte in einem Datensatz
Skewness(*Daten1; Daten2; ... DatenN*)	Berechnet die Werteprobenschiefe der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
SkewnessP(*Daten1; Daten2; ... DatenN*)	Berechnet die Populationsschiefe der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
STANDARDIZE (*x; Mittelw.; Std.Abw.*)	Berechnet einen normalisierten Verteilungswert mithilfe des angegebenen Mittelwert und der genannten Standardabweichung
StdDev(*Daten1; Daten2; ... DatenN*)	Berechnet die Werteproben-Standardabweichung der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
StdDevP(*Daten1; Daten2; ... DatenN*)	Berechnet die Populations-Standardabweichung der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
Sum(*Daten1; Daten2; ... DatenN*)	Berechnet die Summe der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
SUMDEVSQ (*Daten1; Daten2; ... DatenN*)	Berechnet die Quadratsumme der Mittelwertsabweichung der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
SumSQ(*Daten1; Daten2; ... DatenN*)	Berechnet die Quadratsumme der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
STUDENT(*Freiheitsgrad; Statistik; Wert*)	Berechnet die Statistik für die eingegebene Student's t-Verteilung
Variance(*Daten1; Daten2; ... DatenN*)	Berechnet die Werteprobenvarianz der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können
VarianceP(*Daten1; Daten2; ... DatenN*)	Berechnet die Populationsvarianz der betreffenden Argumente, die aus Zahlen, Matrizen oder Bereichen bestehen können

Detaillierte Funktionsbeschreibungen

Nachstehend sind die Statistikfunktionen mit den zugehörigen erforderlichen Argumenten aufgeführt.

AUTOCORRELATION

Beschreibung	Durch **Autocorrelation** (*Daten;Verzöger.*) wird die Autokorrelation für die Werte in einem Datensatz berechnet, und zwar unter Verwendung der Anzahl an Verzögerungen (*Verzöger.*). Bei *Daten* handelt es sich um eine Matrix oder einen Datenbereich, für den die Autokorrelation berechnet werden soll. *Verzöger.* stellt die Anzahl der zu verwendenden Verzögerungen dar.
Beispiele	Durch **StatAutocorrelation** (*C1:C100;1*) wird die Autokorrelation für die Daten im Bereich *C1:C100* mit 1 Verzögerung zurückgegeben.
Richtlinien	Anzahl der Verzögerungen muss größer als oder gleich 1 sein.

AVEDEV

Beschreibung	Durch **AVEDEV**(*Daten1;Daten2;...DatenN*) wird die absolute durchschnittliche Abweichung der betreffenden Daten vom entsprechenden Mittelwert berechnet.
Beispiele	**StatAveDev**(*1;2;5*) berechnet z.B. die absolute durchschnittliche Abweichung der Daten *1, 2* und *5* von ihrem jeweiligen Mittelwert.
Richtlinien	Die Argumente für *Daten1, Daten2* und *DatenN* können aus Zahlen, Matrizen oder Bereichen bestehen.

BINOMIAL

Beschreibung	Durch **BINOMIAL**(*N;P;Statistik;Wert*) wird die Statistik für die binomische Verteilung berechnet, für die die Werte *N* und *P* angegeben sind.
Beispiele	**StatBinomial**(*2;0,3,"x zu p";1*) berechnet z.B. den p-Wert für einen x-Wert von 1 in einer binomischen Verteilung, in der N = 2 und P = 0,3 ist.
Richtlinien	*N* ist die Anzahl der Versuche oder Ereignisse und muss eine Ganzzahl von > 0 sein. *P* stellt die Wahrscheinlichkeit dar und muss >= 0 und <= 1 sein. Bei *Statistik* kann es sich um einen Ganzzahlwert zwischen 1 und 12 oder um eine Zeichenfolge handeln, aus der die zurückzugebende Statistik hervorgeht. Weitere Informationen hierüber sind zu Beginn dieses Kapitels unter **Verteilungsfunktionen** zu finden. Der *Wert* muss >= 0 sein (wenn ein x-Wert eingegeben ist) oder >= 0 und <= 1 (wenn ein p-Wert eingegeben ist).

CATEGORYINDICIES

Beschreibung	Durch **CATEGORYINDICES**(*Bereich;Kategoriename*) werden die Indizes für die Zellen abgerufen, die den angegebenen Kategoriename innerhalb des betreffenden Excel-Bereichs enthalten. Es handelt sich hier um eine Matrix-Funktion und die zurückgegebenen Indizes bestehen aus Indexnummern (zwischen 1 und der Gesamtzahl an Zellen im Bereich), woraus die Positionen der Zellen hervorgehen, die innerhalb des *Bereichs* den *Kategorienamen* enthalten.
Beispiele	**StatCategoryIndices**(*C1:C100;"Männlich"*) gibt z.B. die Indizes (zwischen 1 und 100) der Zellen zurück, die die Zeichenfolge „Männlich“ enthalten.
Richtlinien	*Bereich* ist ein gültiger Excel-Bereich. Bei *Kategoriename* handelt es sich um eine Zeichenfolge, einen Wert oder einen Zellverweis, aus dem die zu suchende Kategorie hervorgeht.

CATEGORYNAMES

Beschreibung	Durch **CATEGORYNAMES**(*Bereich*) werden die Namen der Kategorien im angegebenen Excel-Bereich abgerufen. Es handelt sich hier um eine Matrix-Funktion und die Anzahl der zurückgegebenen Namen liegt zwischen 1 und der Gesamtzahl der im *Bereich* befindlichen Zellen.
Beispiele	**StatCategoryNames**(C1:C100) gibt z.B. die Namen der Kategorien im Bereich C1:C100 zurück.
Richtlinien	*Bereich* ist ein gültiger Excel-Bereich.

CATEGORYOCCURRENCECOUNT

Beschreibung	Durch **CATEGORYOCCURRENCECOUNT**(*Bereich;Kategoriename*) wird die Anzahl der Zellen zurückgegeben, die innerhalb des betreffenden Excel-Bereichs den angegebenen Kategorienamen enthalten.
Beispiele	**StatCategoryOccurrenceCount**(C1:C100,"Männlich") gibt z.B. im Bereich C1:C100 die Anzahl der Zellen zurück, die die Zeichenfolge "Männlich" enthalten.
Richtlinien	*Bereich* ist ein gültiger Excel-Bereich. Bei *Kategoriename* handelt es sich um eine Zeichenfolge, einen Wert oder einen Zellverweis, aus dem die zu suchende Kategorie hervorgeht.

ChiSQ

Beschreibung	Durch **CHISQ**(*Freiheitsgrad;Statistik;Wert*) wird die Statistik für die „schiefe“Chi-Quadrat-Verteilung berechnet, und zwar unter Verwendung der angegebenen Freiheitsgrade.
Beispiele	**StatChiDist**(*2;"x zu p";5*) berechnet z.B. die schiefe Chi-Quadrat-Verteilung mit einem Wert von *5* und *2* Freiheitsgraden.
Richtlinien	Die Anzahl der Freiheitsgrade muss im Bereich von 1 – 32767 liegen. Bei *Statistik* kann es sich um einen Ganzzahlwert zwischen 1 und 12 oder um eine Zeichenfolge handeln, aus der die zurückzugebende Statistik hervorgeht. Weitere Informationen hierüber sind zu Beginn dieses Kapitels unter **Verteilungsfunktionen** zu finden. Der *Wert* muss >= 0 sein (wenn ein x-Wert eingegeben ist) oder >= 0 und <= 1 (wenn ein p-Wert eingegeben ist).

CORRELATIONCOEFF

Beschreibung	Durch **CORRELATIONCOEFF**(*Daten1;Daten2;Flag*) wird der Korrelationskoeffizient zwischen den Datensätzen *Daten1 und Daten2* berechnet. Je nach *Flag*-Wert kann der lineare Koeffizient (Pearson) oder Rangkorrelationskoeffizient (Spearman) berechnet werden.
Beispiele	**StatCorrelationCoeff**(*A1:A100;B1:B100*) berechnet z.B. den Korrelationskoeffizienten zwischen zwei Datensätzen, die sich in den Zellen *A1:A100* und *B1:B100* befinden.
Richtlinien	Die Datensätze *Daten1* und *Daten2* müssen aus der gleichen Anzahl an Elementen bestehen. Falls *Flag* ausgelassen wird oder gleich Null ist, wird der lineare Korrelationskoeffizient (Pearson) berechnet. Wenn *Flag* dagegen nicht gleich Null ist, wird der Rangkorrelationskoeffizient (Spearman) berechnet.

COUNT

Beschreibung	Durch **COUNT**(*Daten1;Daten2;...DatenN*) wird die Anzahl der Elemente in *Daten1, Daten2* bis hin zu *DatenN* berechnet, wobei es sich jeweils um Zahlen, Matrizen oder bereiche handeln kann.
Beispiele	**StatCount**(*A1:A100;B1:B100*) berechnet z.B. die Anzahl an Elementen in den beiden Datensätzen, die sich in den Zellen *A1:A100* und *B1:B100* befinden.
Richtlinien	*Daten1;Daten2;…DatenN* bestehen aus 1 bis 30 Argumenten, bei denen es sich um Zahlen, Matrizen oder Bereiche handeln kann."

COUNTCATEGORIES

Beschreibung	Durch **COUNTCATEGORIES**(*Bereich*) wird die Anzahl der Kategorien im angegebenen Excel-Bereich zurückgegeben.
Beispiele	**StatCountCategories**(C1:C100) gibt z.B. die Anzahl der Kategorien im Bereich C1:C100 zurück.
Richtlinien	*Bereich* ist ein gültiger Excel-Bereich.

COUNTCELLSBYTYPE

Beschreibung	Durch **COUNTCELLSBYTYPE** (*Bereich;Typ*) wird die Anzahl der Elemente in dem eingegebenen Bereich berechnet, die vom angegebenen Typ sind.
Beispiele	**StatCountCellByType**(*A1:A100;1*) berechnet z.B. die Anzahl der Elemente in dem Datensatz im Bereich *A1:A100*, bei denen es sich um Zahlen oder Nummern handelt.
Richtlinien	*Bereich* ist ein gültiger Excel-Bereich. *Typ* ist in *1 = numerisch, 2 = nicht leer, 3 = nicht leer und nicht numerisch* und *4 = leer* unterteilt. Hinweis: In StatTools wird eine Zelle, die nur Leerzeichen enthält, als leere Zelle angesehen.

COUNTRANGE

Beschreibung	Durch **COUNTRANGE** (*Bereich;Minimalwert;Maximalwert;einschl.Min.;einschl.Max.*) wird die Anzahl der Werte im Bereich zwischen *Minimalwert* und *Maximalwert* berechnet. Werte, die genau dem *Minimalwert* und *Maximalwert* entsprechen, können mit einbezogen werden, indem *einschl.Min.* und/oder *einschl.Max.* auf WAHR eingestellt wird.
Beispiele	**StatCountRange**(*A1:A100;1;10;WAHR;WAHR*) berechnet z.B. die Anzahl der Werte in dem Datensatz, der sich in den Zellen *A1:A100* befindet und in den Bereich >= *1* und <= *10* fällt.
Richtlinien	*Bereich* ist der Zellenbereich, in dem die Kategorien gezählt werden sollen. *Minimalwert* ist der minimale Wert für den Bereich. *Maximalwert* ist der maximale Wert für den Bereich. *einschl.Min.* ist ein Boolescher Wert, der anzeigt, ob der Minimalwert in die Zählung mit einbezogen werden soll. Die Standardeinstellung ist WAHR. *einschl.Max.* ist ein Boolescher Wert, der anzeigt, ob der Maximalwert in die Zählung mit einbezogen werden soll. Die Standardeinstellung ist WAHR.

COVARIANCE

Beschreibung	Durch **COVARIANCE**(*Daten1;Daten2*) wird die Werteprobenvarianz zwischen den Datensätzen *Daten1* und *Daten2* berechnet.
Beispiele	**StatCovariance**(*A1:A100;B1:B100*) berechnet z.B. die Werteprobenvarianz zwischen zwei Datensätzen, die sich in den Bereichen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* kann es sich um Matrizen oder Bereiche handeln. Die Datensätze *Daten1* und *Daten2* müssen aus der gleichen Anzahl an Elementen bestehen.

COVARIANCEP

Beschreibung	Durch **COVARIANCEP**(*Daten1;Daten2*) wird die Populationsvarianz zwischen den Datensätzen *Daten1* und *Daten2* berechnet.
Beispiele	**StatCovarianceP**(*A1:A100;B1:B100*) berechnet z.B. die Populationsvarianz zwischen zwei Datensätzen, die sich in den Zellen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* kann es sich um Matrizen oder Bereiche handeln.

DESTACK

Beschreibung	Durch **DESTACK**(*Datenbereich;Kategorienbereich1;Kategorie1;Kategorienbereich2;Kategorie2*) werden die Daten in der angegebenen *Kategorie1* aus den gestapelten Daten im *Datenbereich* extrahiert. Bei dieser Funktion ist es möglich, dass andere StatTools-Statistikfunktionen mit gestapelten Eingabedaten arbeiten. Die Funktion *StatDestack* ist daher immer in andere Funktionen eingebettet, wie hier im Beispiel zu sehen ist. Durch *StatDestack* werden entsprechende Daten aus dem *Datenbereich* extrahiert und in Form einer Matrix an die genannte Kategorie zurückgegeben.
Beispiele	StatMean(**StatDestack**(B1:B100,A1:A100,"Männlich")) berechnet z.B. den Mittelwert für die Werte in dem Bereich B1:B100, in dem der entsprechende Kategoriebereich A1:A100 den Wert "Männlich" zeigt.
Richtlinien	*Datenbereich* ist der Bereich der gestapelten Daten. *Kategorienbereich1* ist der Bereich mit dem ersten Kategorienamen. *Kategorie1* ist die erste Kategorie für den Datenabruf. *Kategorienbereich2 (optional)* ist der Bereich mit dem zweiten Kategorienamen. *Kategorie2 (optional)* ist die zweite Kategorie für den Datenabruf.

DURBINWATSON

Beschreibung	Durch **DURBINWATSON**(*Daten*) wird die Durbin-Watson-Statistik für die Werte im Datensatz *Daten* berechnet.
Beispiele	**StatDurbinWatson**(*A1:A100*) berechnet z.B. die Durbin-Watson-Statistik für den Datenbereich in A1:A100.
Richtlinien	Bei *Daten* kann es sich um eine Matrix oder um einen Datenbereich handeln.

F

Beschreibung	Durch **F**(*Freiheitsgrad1;Freiheitsgrad2;Statistik;Wert*) wird die *Statistik* für die F-Verteilung berechnet, und zwar unter Verwendung des Freiheitsgradzählers *Freiheitsgrad1* und des Freiheitsgradnenners *Freiheitsgrad2.*
Beispiele	**StatF** (1;1;"x zu p";1,5) berechnet z.B. den p-Wert der F-Verteilung für einen x-Wert von *1,5* mit einem Freiheitsgradzählerwert von = *1* und einem Freiheitsgradnennerwert von = *1.*
Richtlinien	*Freiheitsgrad1* und *Freiheitsgrad2* muss jeweils eine Ganzzahl von > 0 sein. Bei *Statistik* kann es sich um einen Ganzzahlwert zwischen 1 und 12 oder um eine Zeichenfolge handeln, aus der die zurückzugebende Statistik hervorgeht. Weitere Informationen hierüber sind zu Beginn dieses Kapitels unter **Verteilungsfunktionen** zu finden. *Wert* muss >= 0 sein.

GETCELLVALUES

Beschreibung	Durch **GETCELLVALUES** (*Bereich,Zelltyp*) werden die Werte für alle Zellen eines bestimmten Typs aus dem Bereich abgerufen.
Beispiele	**StatGetCellValues**(*A1:A100;2*) ruft z.B. die Werte für nicht leere Zellen aus dem Bereich in *A1:A100* ab.
Richtlinien	Bereich ist der Zellenbereich, aus dem die Werte abgerufen werden sollen. *Zelltyp* ist wie folgt aufgeschlüsselt: 0 = alle Zellen, 1 = numerische Zellen , 2 = nicht leere Zellen, 3 = nicht leere und nicht numerische Zellen, 4 = leere Zellen.

KURTOSIS

Beschreibung	Durch **KURTOSIS**(*Daten1;Daten2;...DatenN*) wird die Werteprobenwölbung der in der Funktion angegebenen Daten berechnet. Hinweis: Wird *StatKurtosis* für normal verteilte Daten berechnet, wird dadurch der Wert 3 zurückgegeben.
Beispiele	**StatKurtosis**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Werteprobenwölbung für den Datensatz in *A1:A100* sowie die Werte 1, 2, 3 und 2,4..
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

KURTOSISP

Beschreibung	Durch **KURTOSISP**(*Daten1;Daten2;...DatenN*) wird die Populationswölbung der in der Funktion angegebenen Daten berechnet. Hinweis: Wird *StatKurtosisP* für normal verteilte Daten berechnet, wird dadurch der Wert 3 zurückgegeben.
Beispiele	**StatKurtosisP**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Populationswölbung für den Datensatz in *A1:A100* sowie die Werte 1, 2, 3 und 2,4..
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

LN

Beschreibung	Durch **LN**(*x*) wird der natürliche Logarithmus berechnet.
Beispiele	**StatLN**(*4;5)* berechnet z.B. einen natürlichen Logarithmus von 4,5.
Richtlinien	*x* muss eine positive Realzahl sein.

MAX

Beschreibung	Durch **MAX**(*Daten1;Daten2;...DatenN*) wird das Maximum der Daten berechnet, die durch *Daten1;Daten2;...DatenN* angegeben sind.
Beispiele	**StatMax**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. den Maximalwert für den Datensatz in *A1:A100* sowie für die Werte 1, 2, 3 und 2,4..
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

MEAN

Beschreibung	Durch **MEAN**(*Daten1;Daten2;...DatenN*) wird der Mittelwert der Daten berechnet, die durch *Daten1;Daten2;...DatenN* angegeben sind.
Beispiele	**StatMean**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. den Mittelwert für den Datensatz in *A1:A100* sowie für die Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

MEANABS

Beschreibung	Durch **MEANABS**(*Daten1;Daten2;...DatenN*) wird der Mittelwert des absoluten Werts der Daten berechnet, die durch *Daten1;Daten2;...DatenN* angegeben sind.
Beispiele	**StatMeanAbs**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. den Mittelwert der absoluten Werte für den Datensatz in *A1:A100* sowie für die Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

MEDIAN

Beschreibung	Durch **MEDIAN***(Daten;Kalk.Flag*) wird der Medianwert der Werte in *Daten* berechnet. Dieser Medianwert kann mithilfe von 5 verschiedenen Methoden berechnet werden, die optional durch *Kalk.Flag* angegeben sind.
Beispiele	**StatMedian**(*A1:A100;1*) berechnet z.B. den Medianwert für den Datensatz im Bereich *A1:A100.* Diese Daten sind kontinuierlich.
Richtlinien	Bei *Daten* handelt es sich um einen Excel-Bereich. *Kalk.Flag* ist ein optionales Argument, für das ein Ganzzahlwert im Bereich von -1 bis 5 (siehe nachstehende Auflistung) verwendet werden kann. Dieser Wert entspricht der gewünschten Perzentilberechnungsmethode. -1 oder nicht angegeben – automatisch (auf Basis der Eingabedaten) 0 – entspricht der Perzentilfunktion in Excel (kontinuierlich) 1 – mit asymmetrischen Endpunkten interpoliert (kontinuierlich) 2 – nächstbeste Beobachtung (diskontinuierlich) 3 – erfahrungsgemäße Verteilungsfunktion (diskontinuierlich) 4 – mit symmetrischen Endpunkten interpoliert (kontinuierlich) 5 – erfahrungsgemäße Verteilungsfunktion mit Mittelwertbildung (diskontinuierlich)

MIN

Beschreibung	Durch **MIN**(*Daten1;Daten2;...DatenN*) wird das Minimum der Daten berechnet, die durch *Daten1;Daten2;...DatenN* angegeben sind.
Beispiele	**StatMin**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. den Mindestwert für den Datensatz in *A1:A100* sowie für die Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

MODE

Beschreibung	Durch **MODE**(*Daten;diskont.*) wird der Modus eines Datensatzes berechnet.
Beispiele	**StatMode**(*A1:A100;FALSCH*) berechnet z.B. den Modus des Datensatzes in *A1:A100.* Diese Daten sind kontinuierlich.
Richtlinien	*Daten* ist die Matrix oder der Datenbereich, für den der Modus berechnet werden soll. *diskont*. ist ein optionales Argument, durch das angegeben wird, ob die Daten als diskontinuierlich (wahr) oder kontinuierlich (falsch) behandelt werden sollen. Wenn dieses Argument nicht angegeben ist, wird es automatisch aufgrund der Art von vorhandenen Daten bestimmt."

NORMAL

Beschreibung	Durch **NORMAL**(*Mittelw.;Std.Abw.;Statistik;Wert*) wird die *Statistik* für die Normalverteilung berechnet, die durch *Mittelw*. und *Std.Abw*. angegeben ist.
Beispiele	**StatNormal**(*2;1;"x zu p";3*) berechnet z.B. den p-Wert für einen x-Wert von 3 in einer Normalverteilung, in der der Mittelwert =2 und die Standardabweichung =1 ist.
Richtlinien	*Mittelw*. ist der arithmetische Mittelwert der Verteilung. *Std.Abw*. ist die Standardabweichung der Verteilung und muss > 0 sein. Bei *Statistik* kann es sich um einen Ganzzahlwert zwischen 1 und 12 oder um eine Zeichenfolge handeln, aus der die zurückzugebende Statistik zu erkennen ist. Weitere Informationen hierüber sind zu Beginn dieses Kapitels unter **Verteilungsfunktionen** zu finden. Der *Wert* muss >= 0 und <= 1 sein, wenn ein p-Wert eingegeben wird.

PAIRCOUNT

Beschreibung	Durch **PAIRCOUNT**(*Daten1;Daten2*) wird die Anzahl der Zellenpaare in *Daten1* und *Daten2* gezählt. Es können nur numerische Zellenpaare gezählt werden. Fehlende Werte werden in den Bereichen einfach ignoriert.
Beispiele	**StatPairCount** (*A1:A100;B1:B100*) berechnet z.B. die Anzahl der numerischen Zellenpaare in den beiden Datensätzen, die sich in den Bereichen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* muss es sich um gleich große Excel-Bereiche handeln. In *Daten1* und *Daten2* werden die Zellenpaare zeilenweise ausgewählt, und zwar angefangen von oben links.

PAIRMEAN

Beschreibung	Durch **PAIRMEAN**(*Daten1;Daten2*) wird der Mittelwert der Differenzen zwischen den Zellenpaaren in *Daten1* und *Daten2* berechnet. Es können hierbei nur numerische Zellenpaare berechnet werden. Fehlende Werte in den Bereichen werden einfach ignoriert.
Beispiele	**StatPairMean** (*A1:A100;B1:B100*) berechnet z.B. den Mittelwert zwischen den numerischen Zellenpaaren in den beiden Datensätzen, die sich in den Bereichen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* muss es sich um gleich große Excel-Bereiche handeln. In *Daten1* und *Daten2* werden die Zellenpaare zeilenweise ausgewählt, und zwar angefangen von oben links.

PAIRMEDIAN

Beschreibung	Durch **PAIRMEDIAN**(*Daten1;Daten2;Meth.Flag*) wird der Medianwert der Differenzen zwischen den Zellenpaaren in *Daten1* und *Daten2* berechnet. Es können hierbei nur numerische Zellenpaare berechnet werden. Fehlende Werte in den Bereichen werden einfach ignoriert. Dieser Medianwert kann mithilfe von 5 verschiedenen Methoden berechnet werden, die optional durch *Meth.Flag* angegeben sind.
Beispiele	**StatPairMedian** (*A1:A100;B1:B100*) berechnet z.B. den Medianwert der Differenzen zwischen den numerischen Zellenpaaren in den beiden Datensätzen, die sich in den Bereichen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* muss es sich um gleich große Excel-Bereiche handeln. In *Daten1* und *Daten2* werden die Zellenpaare zeilenweise ausgewählt, und zwar angefangen von oben links. *Meth.Flag* ist ein optionales Argument, für das ein Ganzzahlwert im Bereich von -1 bis 5 verwendet werden kann. Dieser Wert entspricht der gewünschten Berechnungsmethode für den Medianwert. -1 oder nicht angegeben – automatisch (auf Basis der Eingabedaten) 0 – entspricht der Perzentilfunktion in Excel (kontinuierlich) 1 – mit asymmetrischen Endpunkten interpoliert (kontinuierlich) 2 – nächstbeste Beobachtung (diskontinuierlich) 3 – erfahrungsgemäße Verteilungsfunktion (diskontinuierlich) 4 – mit symmetrischen Endpunkten interpoliert (kontinuierlich) 5 – erfahrungsgemäße Verteilungsfunktion mit Mittelwertbildung (diskontinuierlich)

PAIRSTDDEV

Beschreibung	Durch **PAIRSTDDEV**(*Daten1;Daten2*) wird die Werteproben-Standardabweichung der Differenzen zwischen den Zellenpaaren in *Daten1* und *Daten2* berechnet. Es können hierbei nur numerische Zellenpaare berechnet werden. Fehlende Werte in den Bereichen werden einfach ignoriert.
Beispiele	**StatPairStdDev** (*A1:A100;B1:B100*) berechnet z.B. die Werteproben-Standardabweichung der Differenzen zwischen den numerischen Zellenpaaren in den beiden Datensätzen, die sich in den Bereichen *A1:A100* und *B1:B100* befinden.
Richtlinien	Bei *Daten1* und *Daten2* muss es sich um gleich große Excel-Bereiche handeln. In *Daten1* und *Daten2* werden die Zellenpaare zeilenweise ausgewählt, und zwar angefangen von oben links.

PERCENTILE

Beschreibung	Durch **PERCENTILE** (*Daten;p;Meth.Flag*) wird das *p*.-Perzentil der *Daten* berechnet. Perzentile können mithilfe von 5 verschiedenen Methoden berechnet werden, die optional durch *Meth.Flag* angegeben sind.
Beispiele	**StatPercentile**(*A1:A100;0,15;0*) berechnet z.B. das 15. Perzentil für die Daten im Bereich *A1:A100.* Diese Daten sind kontinuierlich und es wird die Perzentilberechnungsmethode mit gewichtetem Durchschnitt verwendet.
Richtlinien	*p* muss 0 – 1 (inklusive) sein. *Meth.Flag* ist ein optionales Argument, für das ein Ganzzahlwert im Bereich von -1 bis 5 verwendet werden kann. Dieser Wert entspricht der gewünschten Perzentilberechnungsmethode. -1 oder nicht angegeben – automatisch (auf Basis der Eingabedaten) 0 – entspricht der Perzentilfunktion in Excel (kontinuierlich) 1 – mit asymmetrischen Endpunkten interpoliert (kontinuierlich) 2 – nächstbeste Beobachtung (diskontinuierlich) 3 – erfahrungsgemäße Verteilungsfunktion (diskontinuierlich) 4 – mit symmetrischen Endpunkten interpoliert (kontinuierlich) 5 – erfahrungsgemäße Verteilungsfunktion mit Mittelwertbildung (diskontinuierlich)

PRODUCT

Beschreibung	Durch **PRODUCT**(*Daten1;Daten2;...DatenN*) wird das Produkt der Daten berechnet, die durch *Daten1;Daten2;...DatenN* angegeben sind.
Beispiele	**StatProduct**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. das Produkt für den Datensatz in *A1:A100* sowie für die Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

QUARTILE

Beschreibung	Durch **QUARTILE** (*Daten;q;Meth.Flag*) wird das angegebene Quartil der *Daten* berechnet. Quartile können mithilfe von 5 verschiedenen Methoden berechnet werden, die optional durch *Meth.Flag* angegeben werden.
Beispiele	**StatQuartile** (*A1:A100;1;FALSCH*) berechnet z.B. das 1. Quartil für die Daten im Bereich *A1:A100.* Diese Daten sind kontinuierlich.
Richtlinien	Bei den *Daten* muss es sich um einen Excel-Bereich handeln. *Q* = Quartil; 0 = Minimum, 1 = 1. Quartil, 2 = 2. Quartil (Medianwert), 3 = 3. Quartil, 4 = Maximum. *Meth.Flag* ist ein optionales Argument, für das ein Ganzzahlwert im Bereich von -1 bis 5 verwendet werden kann. Dieser Wert entspricht der gewünschten Perzentilberechnungsmethode. -1 oder nicht angegeben – automatisch (auf Basis der Eingabedaten) 0 – entspricht der Perzentilfunktion in Excel (kontinuierlich) 1 – mit asymmetrischen Endpunkten interpoliert (kontinuierlich) 2 – nächstbeste Beobachtung (diskontinuierlich) 3 – erfahrungsgemäße Verteilungsfunktion (diskontinuierlich) 4 – mit symmetrischen Endpunkten interpoliert (kontinuierlich) 5 – erfahrungsgemäße Verteilungsfunktion mit Mittelwertbildung (diskontinuierlich)

RAND

Beschreibung	Durch **RAND**() wird ein Zufallswert im Bereich von 0 bis 1 zurückgegeben. In dieser Funktion wird die Zufallswertgenerierung aus @RISK und nicht die in Excel integrierte Zufallswertgenerierung verwendet.
Beispiele	**StatRand**() gibt z.B. einen Zufallswert im Bereich von 0 – 1 zurück

RANGE

Beschreibung	Durch **RANGE**(*Daten1;Daten2;...DatenN*) wird der Bereich (Maximum – Minimum) der Daten berechnet, die durch *Daten1, Daten2* und *DatenN* angegeben sind.
Beispiele	**StatRange**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. den Min-Max-Bereich der Daten in *A1:A100* sowie die Werte *1, 2, 3* und *2,4.*.
Richtlinien	Bei *Daten1;Daten2;…DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

RUNSTEST

Beschreibung	Durch **RUNSTEST**(*Daten;Trennwert*) wird die Ausführzählungsstatistik für die durch *Daten* angegebenen Werte berechnet, und zwar unter Verwendung des *Trennwerts*.
Beispiele	**StatRunsTest**(*A1:A100;StatMean(A1:A100)*) berechnet z.B. die Ausführzählungsstatistik für die Daten in *A1:A100* und verwendet dabei den Mittelwert der Daten als den Trennwert.
Richtlinien	Bei den *Daten* muss es sich um einen Excel-Bereich handeln.

SKEWNESS

Beschreibung	Durch **SKEWNESS**(*Daten1;Daten2;...DatenN*) wird die Werteprobenschiefe in den durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatSkewness**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Werteprobenschiefe für alle Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie für die Werte *1, 2, 3* und *2,4.*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

SKEWNESSP

Beschreibung	Durch **SKEWNESSP**(*Daten1;Daten2;...DatenN*) wird die Populationsschiefe in den durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatSkewnessP**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Populationsschiefe für alle Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie für die Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

STANDARDIZE

Beschreibung	Durch **STANDARDIZE**(*x;Mittelw.;Std.Abw.*) wird ein normalisierter Verteilungswert mithilfe des angegebenen Mittelwert und der angegebenen Standardabweichung berechnet. *x* ist dabei der zu normalisierende Wert.
Beispiele	**StatStandardize**(*2;1;3*) berechnet z.B. einen normalisierten Wert *2* aus einer Verteilung mit einem Mittelwert von *1* und einer Standardabweichung von *3*.
Richtlinien	*x* ist der zu normalisierende Wert. *Mittelw.* ist der arithmetische Mittelwert der Verteilung. *Std.Abw.* ist die Standardabweichung der Verteilung und muss > 0 sein.

STDDEV

Beschreibung	Durch **STDDEV**(*Daten1;Daten2;...DatenN*) wird die Probenstandardabweichung der durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatStdDev**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Probenstandardabweichung für alle Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch für die Werte *1, 2, 3* und *2,4*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

STDDEVP

Beschreibung	Durch **STDDEVP**(*Daten1;Daten2;...DatenN*) wird die Populations-Standardabweichung der durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatStdDevP**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Populations-Standardabweichung für alle Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch für die Werte *1, 2, 3* und *2,4*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

SUM

Beschreibung	Durch **SUM**(*Daten1;Daten2;...DatenN*) wird die Summe der Daten berechnet, die durch *Daten1, Daten2 und DatenN* angegeben sind.
Beispiele	**StatSum**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Summe aller Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch der Werte *1, 2, 3* und *2,4*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

SUMDEVSQ

Beschreibung	Durch **SUMDEVSQ**(*Daten1;Daten2;...DatenN*) wird die Quadratsumme der Mittelwertsabweichung der betreffenden Argumente berechnet, die aus Zahlen, Matrizen oder Bereichen bestehen können. Falls irgendwelche Zahlen fehlen, wird ein Leerwert zurückgegeben.
Beispiele	**StatSumDevSq**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Quadratsumme der Mittelwertabweichung aller Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch der Werte *1, 2, 3* und *2,4*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

SUMSQ

Beschreibung	Durch **SUMSQ**(*Daten1;Daten2;...DatenN*) wird die Quadratsumme der betreffenden Argumente berechnet, die aus Zahlen, Matrizen oder Bereichen bestehen können. Falls irgendwelche Zahlen fehlen, wird ein Leerwert zurückgegeben.
Beispiele	**StatSumSq**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Quadratsumme aller Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch der Werte *1, 2, 3* und *2,4*.
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

STUDENT

Beschreibung	Durch **STUDENT**(*Freiheitsgrad;Statistik;Wert*) wird die *Statistik* für die eingegebene **Student's t-Verteilung** berechnet.
Beispiele	**StatStudent**(5*;1;"x zu p";2*) berechnet z.B. den *p*-Wert aus der Student's t-Verteilung mit *5* Freiheitsgraden bei einem *x*-Wert von *2*.
Richtlinien	*Freiheitsgrad* ist eine Ganzzahl, aus der die Anzahl der Freiheitsgrade hervorgeht. Dieser Wert muss im Bereich von 1 – 32767 liegen. Bei *Statistik* kann es sich um einen Ganzzahlwert zwischen 1 und 12 oder um eine Zeichenfolge handeln, aus der die zurückzugebende Statistik zu erkennen ist. Weitere Informationen hierüber sind zu Beginn dieses Kapitels unter **Verteilungsfunktionen** zu finden. *x* ist der numerische Wert, bei dem die Verteilung ausgewertet wird. Dieser Wert muss >= 0 sein.

VARIANCE

Beschreibung	Durch **VARIANCE**(*Daten1;Daten2;...DatenN*) wird die Werteprobenvarianz der durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatVariance**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Werteprobenvarianz aller Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch der Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

VARIANCEP

Beschreibung	Durch **VARIANCEP**(*Daten1;Daten2;...DatenN*) wird die Werteprobenvarianz der durch *Daten1, Daten2* und *DatenN* angegebenen Daten berechnet.
Beispiele	**StatVarianceP**(*A1:A100;{1;2;3;2,4}*) berechnet z.B. die Werteprobenvarianz aller Werte in dem Datensatz, der sich im Bereich *A1:A100* befindet, sowie auch der Werte *1, 2, 3* und *2,4*..
Richtlinien	Bei *Daten1, Daten2* und *DatenN* kann es sich um Zahlen, Matrizen oder Bereiche handeln.

Index

Anwendungseinstellungen, Befehl, 251
Ausführzählungstest für Zufallsfaktoren, Befehl, 174
Autokorrelation, Befehl, 172
Autorisation, 259
Autorisation, Befehl, 259
Box-Whisker-Plot, Befehl, 131
c-Diagramm, Befehl, 215
Chi-Quadrat-Normungstest, Befehl, 161
Chi-Quadrat-Unabhängigkeitstest, Befehl, 157
Daten transformieren, Befehl, 105
Datensätze löschen, Befehl, 258
Datensatzmanager, Befehl, 95
Deinstallation von StatTools, 7
Dialogfeldspeicher löschen, Befehl, 258
Differenz, Befehl, 109
Diskriminanzanalyse, Befehl, 198
Einseitige ANOVA, Befehl, 150
Histogramm, Befehl, 125
Hypothesetest für Mittelwert/Std. Abweichung, Befehl, 141
Hypothesetest für Proportion, Befehl, 144
Info über, Befehl, 259
Installationsanleitung, 7
Interaktion, Befehl, 111
Kombination, Befehl, 113
Korrelationen und Kovarianz, Befehl, 122
Lilliefors-Test, Befehl, 165
Logistische Regression, Befehl, 193
Mann-Whitney-Test, Befehl, 231
Menüs
  Dienstprogramme, Menü, 101, 251
  Hilfemenü (Modellfenster), 259
  Normalitätstests, Menü, 161
  Regression und Klassifizierung, Menü, 181
  Statistischer Schluss, Menü, 135
  Übersichtsdiagramme, Menü, 125
  Übersichtsstatistik, Menü, 119
  Zeitserie und Prognose, Menü, 169, 203, 223, 239
Normales Q-Q-Diagramm, Befehl, 167
Palisade Corporation, 4
Pareto-Diagramm, Befehl, 204
p-Diagramm, Befehl, 211
Probegrößenauswahl, Befehl, 148
Prognose, Befehl, 176
Pseudo, Befehl, 115
Punktdiagramm, Befehl, 128
StatTools Menüs
  Multivariate Analyse, 239
StatTools-Add-In entladen, Befehl, 258
StatTools-Befehle
  Clusteranalyse, 243
  Hauptkomponenten-Analyse, 239
  Kruskal-Wallis-Test, 235
StatTools-Funktionen, 261
  alphabetische Liste, 267
  StatAutocorrelation, 270
  StatAveDev, 270
  StatBinomial, 271
  StatCategoryIndices, 271
  StatCategoryNames, 272
  StatCategoryOccurrenceCount, 272
  StatChiDist, 272
  StatCorrelationCoeff, 273

StatCount, 273
StatCountCategories, 273
StatCountCellByType, 274
StatCountRange, 274
StatCovariance, 275
StatCovarianceP, 275
StatDestack, 276
StatDurbinWatson, 276
StatF, 277
StatGetCellValues, 277
StatKurtosis, 277
StatKurtosisP, 278
StatLN, 278
StatMax, 278
StatMean, 278
StatMeanAbs, 279
StatMedian, 279
StatMin, 280
StatNormal, 280
StatPairCount, 281
StatPairMean, 281
StatPairMedian, 282
StatPairStdDev, 283
StatPercentile, 283
StatProduct, 284
StatQuartile, 284
StatRand, 285
StatRange, 285
StatRunsTest, 285
StatSkewness, 285
StatSkewnessP, 286
StatStandardize, 286
StatStdDev, 286
StatStdDevP, 287
StatStudent, 288
StatSum, 287
StatSumDevSq, 287
StatSumSq, 288
StatVariance, 288
StatVarianceP, 289
Symbole
Desktop, 7
StatTools, 25
Symbolleisten
StatTools, 25
Systemanforderungen, 6
Übersicht (eine Variable), Befehl, 119
u-Diagramm, Befehl, 218
Variablen entstapeln, Befehl, 104
Variablen stapeln, Befehl, 101
Vertrauensintervall – Mittelwert/Std. Abweichung, Befehl, 135
Vertrauensintervall für Proportionen, Befehl, 138
Verzögerung, Befehl, 107
Vorzeichentest, Befehl, 225
Wilcoxon-Vorzeichen-Rangtest, Befehl, 228
x/r-Diagramme, Befehl, 207
Zeitserien-Diagramm, Befehl, 169
Zufallsprobe, Befehl, 117
Zweiseitige ANOVA, Befehl, 154